

# **EXHIBIT C**

## **Part 3**

Office Action Summary

Application No. 09/597,975	Applicant(s) KONIG ET AL.	
Examiner Barbara N Burgess	Art Unit 2157	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1)  Responsive to communication(s) filed on 08 September 2003.
- 2a)  This action is FINAL.                      2b)  This action is non-final.
- 3)  Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4)  Claim(s) 1-62 is/are pending in the application.
  - 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5)  Claim(s) \_\_\_\_\_ is/are allowed.
- 6)  Claim(s) 1-62 is/are rejected.
- 7)  Claim(s) \_\_\_\_\_ is/are objected to.
- 8)  Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

Application Papers

- 9)  The specification is objected to by the Examiner.
- 10)  The drawing(s) filed on \_\_\_\_\_ is/are: a)  accepted or b)  objected to by the Examiner.
 

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11)  The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. §§ 119 and 120

- 12)  Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a)  All    b)  Some \*    c)  None of:
    - 1.  Certified copies of the priority documents have been received.
    - 2.  Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    - 3.  Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.
- 13)  Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.
  - a)  The translation of the foreign language provisional application has been received.
- 14)  Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

Attachment(s)

- 1)  Notice of References Cited (PTO-892)
- 2)  Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3)  Information Disclosure Statement(s) (PTO-1449) Paper No(s) \_\_\_\_\_
- 4)  Interview Summary (PTO-413) Paper No(s). \_\_\_\_\_
- 5)  Notice of Informal Patent Application (PTO-152)
- 6)  Other: \_\_\_\_\_

Art Unit: 2157

### DETAILED ACTION

This Office Action is in response to amendments filed September 8, 2003. Claims 1-62 are presented for further examination.

#### ***Claim Rejections - 35 USC § 102***

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

2. Claims 1-62 are rejected under 35 U.S.C. 102(e) as being anticipated by Breese et al. (hereinafter "Breese", 6,006,218).

As per claims 1 and 32, Breese discloses a computer-implemented method for providing automatic, personalized information services to a user u, the method comprising:

- Transparently monitoring user interactions with data while the user is engaged in normal use of a computer (column 5, lines 25-38);

Art Unit: 2157

- Updating user-specific data files, wherein the user-specific data files comprise the monitored user interactions with the data and a set of documents associated with the user (column 8, lines 33-36, 44-46);
- Estimating parameters of a learning machine, wherein the parameters define a User Model specific to the user and wherein the parameters are estimated in part from the user-specific data files (column 4 lines 62-67, column 5, lines 1-5, 20-33, column 8, lines 20-26);
- Analyzing a document  $d$  to identify properties of the document (column 8, lines 15-26);
- Estimating a probability  $P(u/d)$  that the document  $d$  is of interest to the user  $u$ , wherein the the probability  $P(u/d)$  is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model (column 6, lines 38-43, column 7, lines 1-10, 31-36, column 9, lines 11-15, 51-55);
- Using the estimated probability to provide automatic, personalized information services to the user (column 16, lines 34-42).

As per claims 2 and 33, Breese discloses wherein the user-specific data files include documents of interest to the user  $u$  and documents that are not of interest to the user  $u$ , and wherein estimating the parameters comprises distinct treatment of the documents of interest and the documents that are not of interest (column 12, lines 44-55).

Art Unit: 2157

As per claims 3 and 34, Breese discloses wherein analyzing the document d provides for the analysis of documents having multiple distinct media types (column 8, lines 15-26)

As per claims 4 and 35, Breese discloses wherein transparently monitoring user interactions with data comprises monitoring multiple distinct modes of user interaction with network data (column 5, lines 25-38).

As per claims 5 and 36, Breese discloses wherein the multiple distinct modes of user interaction comprise a mode selected from the group consisting of a network searching mode, a network navigation mode, a network browsing mode, an email reading mode, and email writing mode, a document writing mode, a viewing "pushed" information mode, a finding expert advice mode, and a product purchasing mode (column 5, lines 25-38).

As per claims 6 and 37, Breese discloses crawling network documents, wherein the crawling comprises parsing crawled documents for links, calculating probable user interest in the parsed links using the learning machine, and preferentially following links likely to be of interest to the user (column 9, lines 51-67, column 10, lines 1-27, 38-55).

Art Unit: 2157

As per claims 7 and 38, Breese discloses wherein the identified properties of the document  $d$  comprise a user  $u$ -independent property selected from the group consisting of:

- A probability  $P(t|d)$  that the document  $d$  is of interest to users interested in a topic  $t$  (column 6, lines 38-45);
- A topic classifier discrete probability distribution  $P(t/d)$  (column 6, lines 38-45);
- A product model discrete probability distribution  $P(p/d)$  (column 6, lines 38-45);
- Product feature values extracted from the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- An author of the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- An age of the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- A list of documents linked to the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- A language of the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- A number of users who have accessed the document  $d$  (column 11, lines 1-30);
- A number of users who have saved the document  $d$  in a favorite document list (column 11, lines 1-30);
- A list of users previously interested in the document  $d$  (column 11, lines 1-30).

As per claims 8 and 39, Breese discloses wherein the parameters of the learning machine define a user  $u$ -dependent function selected from the group consisting of:

Art Unit: 2157

- A user topic probability distribution  $P(t/u)$  representing interests of the user  $u$  in various topics  $t$  (column 6, lines 38-45);
- A user product probability distribution  $P(p/u)$  representing interests of the user  $u$  in various products  $p$  (column 6, lines 38-45);
- A user product feature probability distribution  $P(F/u, p)$  representing interests of the user  $u$  in various features  $f$  of each of the various products  $p$  (column 6, lines 38-45, column 9, lines 50-67, column 10, lines 1-20);
- A website probability distribution  $P(s/u)$  representing interests of the user  $u$  in various websites  $s$  (column 8, lines 1-27);
- A cluster probability distribution  $P(c(u)/u)$  representing similarity of the user  $u$  to users in various clusters  $c(u)$  (column 9, lines 25-40);
- A phrase model probability distribution  $P(w/u)$  representing interests of the user  $u$  in various phrases  $w$  (column 9, lines 25-40);
- An information theory based measure  $I(lw; lu)$  representing mutual information between various phrases  $w$  and the user  $u$  (column 9, lines 25-40);
- An information theory based measure  $I(It; lu)$  representing mutual information between various topics  $t$  and the user  $u$  (column 6, lines 38-45);
- An information theory based measure  $I(Is; lu)$  representing mutual information between various websites  $s$  and the user  $u$  (column 8, lines 1-27);
- An information theory based measure  $I(lp; lu)$  representing mutual information between various products  $p$  and the user  $u$  (column 8, lines 1-27);

Art Unit: 2157

- An information theory based measure  $I(f; u)$  representing mutual information between various features  $f$  of each of the various products  $p$  and the user  $u$  (column 8, lines 1-27).

As per claims 9 and 40, Breese discloses wherein the parameters of the learning machine define:

- A user product probability distribution  $P(p/u)$  representing interests of the user  $u$  in various products  $p$  (column 8, lines 1-27);
- A user product feature probability distribution  $P(f/u, p)$  representing interests of the user  $u$  in various features  $f$  of each of the various products  $p$  (column 8, lines 1-27);
- Estimating a probability  $P(u/d, \text{product described}=p)$  that a document  $d$  that describes a product  $p$  is of interest to the user  $u$ , wherein the probability is estimated in part from the user product probability distribution and the user product feature probability distribution (column 8, lines 1-27).

As per claims 10 and 41, Breese discloses recommending products to the user based on the probability  $P(u/d, \text{product described}=p)$  (column 6, lines 38-45).

As per claims 11 and 42, Breese discloses estimating a posterior probability  $P(u/d, q)$  that the document  $d$  is of interest to the user  $u$ , given a query  $q$  submitted by the user (column 8, lines 44-67).



Art Unit: 2157

As per claims 12 and 43, Breese discloses wherein estimating the posterior probability comprises estimating a probability  $P(q/d, u)$  that the query  $q$  is expressed by the user  $u$  with an information need in the document  $d$  (column 8, lines 44-67).

As per claims 13 and 44, Breese discloses applying the identified properties of the document  $d$  to a learning machine having product parameters characterizing a product  $p$  to estimate a probability  $P(p/d)$  that the document  $d$  refers to the product  $p$  (column 6, lines 38-45).

As per claims 14 and 45, Breese discloses updating the product parameters based on the identified properties of the document  $d$  and the estimated probability  $P(p/d)$  (column 8, lines 33-36, 44-46).

As per claims 15 and 46, Breese discloses initializing the product parameters based on a set of documents associated with the product  $p$  (column 8, lines 15-50).

As per claims 16 and 47, Breese discloses clustering multiple users into clusters of similar users, wherein the clustering comprises calculating distances between User Models, and selecting similar users based on the calculated distances between User Models (column 9, lines 25-40).

Art Unit: 2157

As per claims 17 and 48, discloses calculating relative entropy values between User Models of multiple users, and clustering together users based on the calculated relative entropy values (column 9, lines 25-40).

As per claims 18 and 49, Breese discloses wherein the parameters defining the User Model comprise calculated distances between the User Model and User Models of users similar to the user (column 11, columns 13-14).

As per claim 19 and 50, Breese discloses selecting in a group of users an expert user in an area expertise, wherein selecting the expert user comprises finding an expert User Model among User Models of the group of users, such that the expert User Model indicates a strong interest of the expert user in a document associated with the area of expertise (column 9, lines 20-50).

As per claim 20 and 51, Breese discloses parsing the document  $d$  for hyperlinks, and separately estimating for each of the hyperlinks a probability that the hyperlink is of interest to the user  $u$  (column 9, lines 51-67, column 10, lines 1-27, 38-55).

As per claims 21 and 52, Breese discloses sending to a third party web server user interest information derived from the User Model, whereby the third party web server may customize its interaction with the user (column 17, lines 6-25).

Art Unit: 2157

As per claims 22 and 53, Breese discloses wherein the monitored user interactions include a sequence of interaction times (column 9, lines 63-67).

As per claims 23 and 54, Breese discloses initializing the User Model using information selected from the group consisting of a set of documents provided by the user, a web browser history file associated with the user, a web browser bookmarks file associated with the user, ratings by the user of a set of documents, and previous product purchases made by the user (column 5, lines 25-38).

As per claims 24 and 55, Breese discloses modifying the User Model based on User Model modification requests provided by the user (column 8, lines 15-60).

As per claims 25 and 56, Breese discloses providing to the user a score for a document identified by the user, wherein the score is derived from the estimated probability (column 12, lines 49-55, column 13, lines 1-5).

As per claims 26 and 57, Breese discloses providing to the user a 3D map of a hyperlinked document collection, wherein the 3D map indicates a user interest in each document (column 5, lines 25-38).

Art Unit: 2157

As per claims 27 and 58, Breese discloses temporarily using a User Model that is built from a set of predetermined parameters of a profile selected by the user (column 8, lines 15-60).

As per claims 28 and 59, Breese discloses initializing the User Model by selecting a set of predetermined parameters of a prototype user selected by the user (column 8, lines 15-60).

As per claims 29 and 60, Breese discloses updating the predetermined parameters of the prototype user based on actions of users similar to the prototype user (column 8, lines 33-36, 44-46).

As per claims 30 and 61, Breese discloses identifying a set of users interest in the document d (column 16, lines 34-42).

As per claims 31 and 62, Breese discloses calculating a range of interests in the document d for the identified set of users (column 16, lines 34-42).

### ***Response to Arguments***

**The Office notes the following arguments:**

Art Unit: 2157

- (a) The important difference between the knowledge probability disclosed by Breese et al. and the probability that a document is of interest to a user in the current application appears to have been overlooked in the Office Action.
- (b) The user's information retrieval request of Breese et al. is not a User Model and the estimate of the value to the user of a reviewing the entry of Breese et al. is not a probability and is too vague to be considered a User Model.
- (c) The cited section of Breese et al. do not discuss any analysis of documents and are irrelevant to the claim element.
- (d) The parameters being estimated in the claim have nothing to do with the probability that the user knows of any specific items.
- (e) The applicant asserts that the cited passages do not specify nor imply that the user is engaged in normal use of the computer, nor that the monitoring is transparent.
- (f) Breese et al. Refer only to knowledge probability, not the probability that the documents are of interest to the user.
- (g) Breese does not disclose any analyzing of documents; hence these claims are novel and nonobvious.
- (h) Breese does not discuss nor suggest products. Therefore, Breese cannot anticipate, nor make obvious claims that include limitations to products.
- (i) Breese discusses neither clusters nor entropy values. Therefore, Breese does not anticipate nor make obvious claims referencing this.
- (j) Breese does not discuss calculated distances. Hence, Breese does not anticipate nor make obvious these claims.

Art Unit: 2157

(k) Breese do not discuss the selection of an expert user, hence these claim are not anticipated nor made obvious by Breese.

(l) Breese does not disclose the parsing of document for hyperlinks.

(m) Breese does not disclose how long the user's interaction time is.

(n) Breese does not disclose User Model modification requests.

(o) Breese does not disclose the use of parameters of a prototype user that is selected by the user.

3. Applicant's arguments filed have been fully considered but they are not persuasive.

**In response to:**

(a), (d), (f), Breese does indeed disclose probability that a document is of interest to the user. This is done based on whether the user already knows of a specific document. If the document is already known to the user, it may then be thought of as unwanted or not useful which merely distracts the user from more useful material and/or wastes the user's time. "The list of recommendations generated using a collaborative filter can be, e.g., a list of television shows, songs, World Wide Web sites, etc. that may be of interest to the user. In accordance with the present invention, the value of individual recommendations, in a list of recommendations, is adjusted, e.g., corrected, based on the estimated probability that the user already knows a particular item being recommended to the user. By performing a re-ranking operation in this manner, the

Art Unit: 2157

present invention achieves a more meaningful ranking of recommendations than is the achieved using prior art approaches. Furthermore, the noise problem and the risk of information overload associated with including known items in the results of an information retrieval request, such as a collaborative filtering operation, is reduced. The present invention can be used when making recommendations to a user, e.g., in response to a user initiated information request, or after monitoring a user's actions for a period of time" (column 3, lines 8-26). However, not only does Breese disclose the probability that a document is of interest to a user, it takes into consideration whether the user knows of the document so that the user has new and useful documents and not stale, outdated, not useful information. Therefore, Breese et al. definitely discloses the probability that the document is of interest to the user.

(b) Breese clearly discloses a User Model specific to the user. Breese discloses a database that has information about the user and the user's interests (column 4 lines 62-67, column 5, lines 1-5, 20-33, column 8, lines 20-26).

(c), (g), (h), (k)-(l), (o) Applicant's arguments fail to comply with 37 CFR 1.111(b) because they amount to a general allegation that the claims define a patentable invention without specifically pointing out how the language of the claims patentably distinguishes them from the references.

Art Unit: 2157

(e) Breese specifically discloses monitoring a user's actions while accessing Internet sites (column 3, lines 23-30). Therefore, Breese undoubtedly teaches the user in normal use and monitoring being transparent to the user.

(i)-(j), (n) Breese discloses a poll on how others view certain documents in respect to the actual user (column 5, lines 51-67, column 6, lines 1-10, 38-46). Therefore, Breese plainly discloses calculated distances between the User Model and the User Models of other users.

(m) Breese plainly discloses the amount of time the user interacts with a document (column 5, lines 20-36, column 8, lines 4-8, 21-25, column 9, lines 65-67).

### ***Conclusion***

4. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.



Art Unit: 2157

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Barbara N Burgess whose telephone number is (703) 305-3366. The examiner can normally be reached on M-F (8:00am-4:00pm).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Ettinene can be reached on (703) 308-7562. The fax phone numbers for the organization where this application or proceeding is assigned are (703) 872-9306 for regular communications and (703) 872-9306 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-3900.

Barbara N Burgess  
Examiner  
Art Unit 2157

\*\*\*

November 26, 2003

  
ARIO ETIENNE  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2100

**Notice of References Cited**

Application/Control No.  
09/597,975

Applicant(s)/Patent Under  
Reexamination  
KONIG ET AL.

Examiner  
Barbara N Burgess

Art Unit  
2157

Page 1 of 1

**U.S. PATENT DOCUMENTS**

* A	Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
A	US-6,006,218	12-1999	Breese et al.	707/3
B	US-			
C	US-			
D	US-			
E	US-			
F	US-			
G	US-			
H	US-			
I	US-			
J	US-			
K	US-			
L	US-			
M	US-			

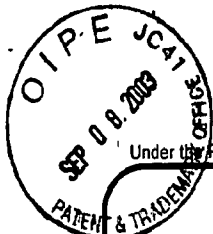
**FOREIGN PATENT DOCUMENTS**

* N	Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
N					
O					
P					
Q					
R					
S					
T					

**NON-PATENT DOCUMENTS**

* U	Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages
U	
V	
W	
X	

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)  
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.



Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

<b>TRANSMITTAL FORM</b> <i>(to be used for all correspondence after initial filing)</i>	<b>Application Number</b>	09/597,975	
	<b>Filing Date</b>	06/20/2000 <b>RECEIVED</b>	
	<b>First Named Inventor</b>	Konig SEP 10 2003	
	<b>Group Art Unit</b>	2157 Technology Center 2100	
	<b>Examiner Name</b>	Burgess	
<b>Total Number of Pages in This Submission</b>	11	<b>Attorney Docket Number</b>	UTO-101

<b>ENCLOSURES (check all that apply)</b>		
<input type="checkbox"/> Fee Transmittal Form	<input type="checkbox"/> Assignment Papers (for an Application)	<input type="checkbox"/> After Allowance Communication to Group
<input type="checkbox"/> Fee Attached	<input type="checkbox"/> Drawing(s)	<input type="checkbox"/> Appeal Communication to Board of Appeals and Interferences
<input checked="" type="checkbox"/> Amendment / Reply	<input type="checkbox"/> Licensing-related Papers	<input type="checkbox"/> Appeal Communication to Group (Appeal Notice, Brief, Reply Brief)
<input type="checkbox"/> After Final	<input type="checkbox"/> Petition	<input type="checkbox"/> Proprietary Information
<input type="checkbox"/> Affidavits/declaration(s)	<input type="checkbox"/> Petition to Convert to a Provisional Application	<input type="checkbox"/> Status Letter
<input type="checkbox"/> Extension of Time Request	<input type="checkbox"/> Power of Attorney, Revocation Change of Correspondence Address	<input type="checkbox"/> Other Enclosure(s) (please identify below):
<input type="checkbox"/> Express Abandonment Request	<input type="checkbox"/> Terminal Disclaimer	
<input type="checkbox"/> Information Disclosure Statement	<input type="checkbox"/> Request for Refund	
<input type="checkbox"/> Certified Copy of Priority Document(s)	<input type="checkbox"/> CD, Number of CD(s) _____	
<input type="checkbox"/> Response to Missing Parts/ Incomplete Application	Remarks	
<input type="checkbox"/> Response to Missing Parts under 37 CFR 1.52 or 1.53		

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT	
Firm or Individual name	Dr. Ron Jacobs Reg. 50,142
Signature	
Date	9-4-03

CERTIFICATE OF MAILING	
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O.Box 1450, Alexandria, VA 22313	
date: 9-4-03	
Typed or printed name	Sarah Treat
Signature	
Date	9-4-03

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

4/A  
RWB  
9-17-03

In re Application of: Konig, Twersky, and Berthold  
Serial No.: 09/597,975                      Examiner: Burgess  
Filed: June 20, 2000                      Art Unit: 2157  
For: Automatic, Personalized Online Information and Product Services

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to:  
Commissioner for Patents,  
P. O. Box 1450  
Alexandria, Virginia 22313-1450

on September 4, 2003.

RECEIVED

SEP 10 2003

Technology Center 2100

Print Name: Sarah Treat

Signature: [Handwritten Signature]

AMENDMENT

Commissioner for Patents,  
P. O. Box 1450  
Alexandria, Virginia 22313-1450

Sir:

In response to an Office Action mailed by the U.S.P.T.O. on June 4, 2003, please amend the above-identified patent application as indicated below.

## AMENDMENTS TO THE SPECIFICATION

Please amend the second paragraph of BACKGROUND ART, starting on page 1,  
to read as follows:

---

A | Information filtering techniques focus on the analysis of item content and the development of a personal user interest profile. In the simplest case, a user is characterized by a set of documents, actions regarding previous documents, and user-defined parameters, and new documents are characterized and compared with the user profile. For example, U.S. Patent No. 5,933,827, issued to Cole et al., discloses a system for identifying new web pages of interest to a user. The user is characterized simply by a set of categories, and new documents are categorized and compared with the user's profile. U.S. Patent No. 5,999,975, issued to Kittaka et al., describes an online information providing scheme that characterizes users and documents by a set of attributes, which are compared and updated ~~base~~ based on user selection of particular documents. U.S. Patent No. 6,006,218, issued to Breese et al., discloses a method for retrieving information based on a user's knowledge, in which the probability that a user already knows of a document is calculated based on user-selected parameters or popularity of the document. U.S. Patent No. 5,754,939, issued to Herz et al., discloses a method for identifying objects of interest to a user based on stored user profiles and target object profiles. Other techniques rate documents using the TFIDF (term frequency, inverse document frequency) measure. The user is represented as a vector of the most informative words in a set of user-associated documents. New documents are parsed to obtain a list of the most informative words, and this list is compared to the user's vector to determine the user's interest in the new document.

---

## REMARKS

Claims 1-62 are pending in the application. Claims 1-62 are rejected. Specifically, claims 1-62 are rejected under 35 U.S.C. 102(e) as being anticipated by Breese et al.

A paragraph in the specification has been amended to replace “base” with “based”.

The Office Action asserts that all of the elements and limitations of all of the claims are found in Breese et al. The applicants respectfully disagree.

Addressing first claims 1 and 32, the Office Action asserts that column 6, lines 38-43, column 7, lines 1-10, 31-36, and column 9, lines 11-15, 51-55 disclose “estimating a probability  $P(u|d)$  that the document  $d$  is of interest to the user  $u$ , wherein the probability  $P(u|d)$  is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model.” In column 6, lines 38-43, Breese, et al. discuss the probability that a user knows about an item, not that a document is of interest to a user. Lines 1-10 and 31-36 of column 7 of Breese, et al. similarly discuss the probability that user knows about an item. Interestingly, in lines 36-38 of column 7, Breese, et al. state, “In the illustrated embodiment, the higher the probability is, that the user already knows of an item, the lower the adjusted score and rank assigned to that item will be.” Hence, in Breese, et al., if an item has a high probability of being known to the user, then its score will be reduced. However, in the current application, the applicants are concerned with whether a document is of interest to a user. Specifically, on page 30 of the current application, the applicants state, “Positive examples are documents of interest to a user: search results that are visited following a search query, documents saved in the user favorites or bookmarks file, web sites that the user visits independently of search queries, etc.” Hence, positive examples of interest as understood in the current application, would be indicators that the user knows of the document and therefore would reduce the score in Breese, et al. Hence the basic

premises and goals of Breese, et al. are not consistent with those of the current application. This important difference between the knowledge probability disclosed by Breese, et al., and the probability that a document is of interest to a user in the current application appears to have been overlooked in the Office Action.

Column 9, lines 11-15 of Breese, et al. state, “The score can correspond to the entry’s estimated, e.g. calculated, probability of relevance to a user’s information retrieval request. Alternatively, the score may be an estimate of the value to the user of reviewing the entry.” Note that the claims specifically require the estimation of a probability, and that this probability be estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model. The current application on page 11 states, “All of the above features of Personal Web 12 are based on a User Model 13 that represents user interests in a document or product independently of any specific user information need, i.e., not related to a specific query.” Hence the “user’s information retrieval request” of Breese, et al. from column 9, lines 11-15, is definitely not a User Model, and the “estimate of the value to the user of a reviewing the entry” of Breese, et al. is not a probability and is too vague to be considered a User Model. Hence, this element of claims 1 and 32 is not found in the cited reference.

The Office Action also asserts that element, “analyzing a document to identify properties of the document,” is described in column 8, lines 15-26 of Breese, et al. The cited section of Breese, et al. do not discuss any analysis of documents and are irrelevant to the claim element.

Additionally, the Office Action asserts that column 7, lines 2-5 of Breese, et al. disclose, “estimating parameters of a learning machine, wherein the parameters define a User Model specific to the user and wherein the parameters are estimated in part from the user-specific data files.” Lines 2-5 only state that a knowledge probability estimator is used to estimate the probability that the user already knows about the items. The parameters being estimated in the claim have nothing to do with the probability that the user knows of any specific items.

Finally, with respect to claims 1 and 32, the Office Action asserts that column 5, lines 25-38 disclose, “transparently monitoring user interactions with data while the user

is engaged in normal use of a computer.” The applicants assert that the cited passages do not specify nor imply that the user is engaged in normal use of the computer, nor that the monitoring is transparent. In fact, the cited passage includes obtaining information from questionnaire results, which are certainly not transparently obtained when the user is engaged in normal use of a computer.

The applicants believe that the significant differences discussed above between the claimed invention and Breese, et al. make the claimed invention novel and nonobvious. Because all other claims depend from either claim 1 or claim 32, the applicants believe that all pending claims are novel and nonobvious.

In addition to their dependency on claims 1 or 32, the applicants believe that the following dependent claims further distinguish from the cited reference.

The Office Action states that Breese, et al. disclose claims 2 and 33 in column 12, lines 44-55. However, Breese, et al. refer only to knowledge probability, not the probability that the documents are of interest to the user. As discussed above, these are very different concepts. Breese, et al. do not disclose nor suggest the claim limitations of claims 2 and 33. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

As per claims 3 and 34, the Office Action states that Breese, et al. disclose limitations relating to the analyzing of documents. However, Breese, et al. do not disclose any analyzing of documents, hence these claims are novel and nonobvious. The Office Action states that claims 6 and 37 are anticipated by Breese, et al. These claims include additional limitations relating to probable user interest. As noted above, Breese, et al. disclose the probability that the user knows of a document, not that the user is interested in a document, let alone documents having multiple distinct media types. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

As per claims 8 and 39, the Office Action states that various limitations of the claims are disclosed in Breese, et al. However, these limitations are actually not disclosed by Breese, et al. In particular, Office Action states that the topic probability distribution is disclosed in column 6, lines 38-45. However, only knowledge probability is discussed there, not the probability that the user will have an interest in the topic. Similarly, the Office Action states that the product probability distribution is disclosed in column 6, lines 38-45. Note that Breese, et al. does not discuss products at all, hence this and all



other limitations of this claim that focus on products are not disclosed by Breese, et al. The Office Action states that the website probability distribution is disclosed in column 8, lines 1-27. However, these lines only disclose knowledge probability, not the probability that the user will have an interest in the website. The Office Action states that the cluster probability distribution is disclosed in column 9, lines 25-40. However, these lines only disclose knowledge probability, not the probability that the user is similar to other users. The Office Action states that a phrase model probability is disclosed in column 9, lines 25-40. However, these lines only disclose knowledge probability, not the probability that the user is interested in various phrases. The Office Action states that an information theory based measure representing mutual information between various phrases is disclosed in column 9, lines 25-40. However, these lines do not disclose anything related to mutual information, as required by the claim limitation. The Office Action states that an information theory based measure representing mutual information between various topics is disclosed in column 6, lines 38-45. However, these lines do not disclose anything related to mutual information, as required by the claim limitation. The Office Action states that an information theory based measure representing mutual information between various websites is disclosed in column 8, lines 1-27. However, these lines do not disclose anything related to mutual information, as required by the claim limitation. The Office Action states that an information theory based measure representing mutual information between various products is disclosed in column 8, lines 1-27. However, these lines do not disclose anything related to mutual information, as required by the claim limitation, nor do they suggest products. The Office Action states that an information theory based measure representing mutual information between various features of products in column 8, lines 1-27. However, these lines do not disclose anything related to mutual information, as required by the claim limitation, nor do they suggest products or the features of products. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

With respect to claims 9-10, and 40-41, note that Breese, et al. do not discuss nor suggest products. Therefore, Breese, et al. cannot anticipate, nor make obvious claims that include limitations to products.

With respect to claims 11-12, and 42-43, these claims include limitations to the posterior probability that a document is of interest to a user given a particular query. Column 8, lines 44-67 of Breese, et al. do not discuss nor make obvious such probabilities.

With respect to claims 14-15 and 45-46, Breese, et al. do not include any discussion nor suggestion of products. Because these claims include limitations directed to products, Breese, et al. do not anticipate, nor make obvious these claims.

With respect to claims 16-17 and 47-48, the cited lines (column 9, lines 25-40) refer to scores "associated with each entry in the list of search results." The claim limitations in 16 and 47 refer to clustering multiple users into clusters of similar users. And the claim limitations in 17 and 48 include limitations associated with the calculation of relative entropy values between User Models of multiple users. Breese, et al. discuss neither clusters nor entropy values. Therefore, Breese, et al. do not anticipate nor make obvious these claims.

As per claims 18 and 49, column 11 discloses estimates that a user knows of an item and columns 13-14 discuss scores relating to knowledge of a particular item. The claim limitations refer to calculated distances between the User Model and the User Models of users similar to the user. Breese, et al. do not discuss such calculated distances, hence Breese, et al. do not anticipate nor make obvious these claims.

With respect to claims 19 and 50, these claims include limitations relating to the selection of an expert user from among a group of users. Breese, et al. do not discuss the selection of an expert user, hence these claims are not anticipated nor made obvious by Breese, et al.

As per claims 20 and 51, Breese, et al. do not disclose the parsing of a document for hyperlinks as specified in the claims. Therefore Breese, et al. do not anticipate nor make obvious these claims.

With respect to claims 21 and 52, column 17, lines 6-25 make no mention of sending to a third party web server interest information derived from the User Model. Note that Breese, et al. do not discuss interest information but knowledge probabilities. As previously discussed these are very different from each other. Therefore, Breese, et al. do not anticipate nor make obvious these claims.

As per claims 22 and 53, column 9, lines 63-67 do include the amount of time a user has been on the Internet as a factor that can be used in generating knowledge probabilities. However, as used in Breese, et al., "the amount of time a user has been on the Internet" is used to indicate the user's experience with the Internet, not how long the user's interaction time is, as specified in the claims. Therefore Breese, et al. do not anticipate nor make obvious these claims.

As per claims 24 and 55, column 8, lines 15-60 of Breese, et al. disclose updating of a user database without the user taking any explicit action. The User Model modification requests in the claims are explicitly requested by the user. Breese, et al. do not disclose such a modification, therefore the claims are not anticipated nor made obvious by the cited reference.

With respect to claims 25 and 56, the claims involve a score derived from the estimated probability. The estimated probability is that of user interest in a document (as specified in claim 1). The cited portions of Breese, et al. discuss scores generated using knowledge probabilities, which as discussed previously, are very different. Therefore Breese, et al. do not anticipate nor make obvious the claims.

As per claims 26 and 57, column 5, lines 25-38 of Breese, et al. discuss information about a user. However, there is no discussion of a 3D map of a hyper linked document collection, nor the user's interest in each document, as specified in the claims. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

With respect to claims 27 and 58, column 8, lines 15-60 discuss the use of personal information. However, the claim specifies that the user temporarily uses a User Model that is built from a set of predetermined parameters of a profile selected by the user. Breese, et al. do not discuss anything like this, hence the claims are not anticipated nor made obvious.

As per claims 28 and 59, Breese, et al. do not disclose the use of parameters of a prototype user that is selected by the user. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

With respect to claims 29 and 60, because Breese, et al. do not discuss a prototype user, they cannot update the parameters of a prototype users. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

As per claims 30 and 61, Breese, et al. only refers to knowledge probabilities, not user interest in documents. In addition, column 16, lines 34-42 make no mention of identifying a set of users interested in a given document. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

As per claims 31 and 62, Breese, et al. includes no discussion as to calculating a range of interests in the document for the identified set of users. Therefore, Breese, et al. do not anticipate nor make obvious the claims.

CONCLUSION

In view of the above comments, the applicants respectfully submit that this patent application is in condition for allowance. Early action to this end is requested.

Respectfully submitted,

A handwritten signature in black ink, appearing to be 'Dr. Ron Jacobs', written in a cursive style.

Dr. Ron Jacobs

Registration Number 50,142



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER OF PATENTS AND TRADEMARKS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/597,975	06/20/2000	Yochai Konig	UTO-101	9014

30869 7590 06/04/2003

LUMEN INTELLECTUAL PROPERTY SERVICES  
2345 YALE STREET  
SUITE 200  
PALO ALTO, CA 94306

EXAMINER

BURGESS, BARBARA N

ART UNIT PAPER NUMBER

2157

DATE MAILED: 06/04/2003

3

Please find below and/or attached an Office communication concerning this application or proceeding.

**Office Action Summary**

Application No.

09/597,975

Applicant(s)

KONIG ET AL.

Examiner

Barbara N Burgess

Art Unit

2157

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1)  Responsive to communication(s) filed on 20 June 2000.
- 2a)  This action is FINAL.
- 2b)  This action is non-final.
- 3)  Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4)  Claim(s) 1-62 is/are pending in the application.
  - 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5)  Claim(s) \_\_\_\_\_ is/are allowed.
- 6)  Claim(s) 1-62 is/are rejected.
- 7)  Claim(s) \_\_\_\_\_ is/are objected to.
- 8)  Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9)  The specification is objected to by the Examiner.
- 10)  The drawing(s) filed on \_\_\_\_\_ is/are: a)  accepted or b)  objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11)  The proposed drawing correction filed on \_\_\_\_\_ is: a)  approved b)  disapproved by the Examiner.  
If approved, corrected drawings are required in reply to this Office action.
- 12)  The oath or declaration is objected to by the Examiner.

**Priority under 35 U.S.C. §§ 119 and 120**

- 13)  Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a)  All b)  Some \* c)  None of:
    - 1.  Certified copies of the priority documents have been received.
    - 2.  Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    - 3.  Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.
- 14)  Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
  - a)  The translation of the foreign language provisional application has been received.
- 15)  Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

**Attachment(s)**

- 1)  Notice of References Cited (PTO-892)
- 2)  Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3)  Information Disclosure Statement(s) (PTO-1449) Paper No(s) 2.
- 4)  Interview Summary (PTO-413) Paper No(s). \_\_\_\_\_.
- 5)  Notice of Informal Patent Application (PTO-152)
- 6)  Other:

Art Unit: 2157

## DETAILED ACTION

### *Claim Rejections - 35 USC § 102*

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

2. Claims 1- are rejected under 35 U.S.C. 102(e) as being anticipated by Breese et al. (hereinafter "Breese", 6,006,218).

As per claims 1 and 32, Breese discloses a computer-implemented method for providing automatic, personalized information services to a user u, the method comprising:

- Transparently monitoring user interactions with data while the user is engaged in normal use of a computer (column 5, lines 25-38);
- Updating user-specific data files, wherein the user-specific data files comprise the monitored user interactions with the data and a set of documents associated with the user (column 8, lines 33-36, 44-46);



Art Unit: 2157

- Estimating parameters of a learning machine, wherein the parameters define a User Model specific to the user and wherein the parameters are estimated in part from the user-specific data files (column 7, lines 2-5);
- Analyzing a document  $d$  to identify properties of the document (column 8, lines 15-26);
- Estimating a probability  $P(u/d)$  that the document  $d$  is of interest to the user  $u$ , wherein the the probability  $P(u/d)$  is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model (column 6, lines 38-43, column 7, lines 1-10, 31-36, column 9, lines 11-15, 51-55);
- Using the estimated probability to provide automatic, personalized information services to the user (column 16, lines 34-42).

As per claims 2 and 33, Breese discloses wherein the user-specific data files include documents of interest to the user  $u$  and documents that are not of interest to the user  $u$ , and wherein estimating the parameters comprises distinct treatment of the documents of interest and the documents that are not of interest (column 12, lines 44-55).

As per claims 3 and 34, Breese discloses wherein analyzing the document  $d$  provides for the analysis of documents having multiple distinct media types (column 8, lines 15-26)

Art Unit: 2157

As per claims 4 and 35, Breese discloses wherein transparently monitoring user interactions with data comprises monitoring multiple distinct modes of user interaction with network data (column 5, lines 25-38).

As per claims 5 and 36, Breese discloses wherein the multiple distinct modes of user interaction comprise a mode selected from the group consisting of a network searching mode, a network navigation mode, a network browsing mode, an email reading mode, and email writing mode, a document writing mode, a viewing "pushed" information mode, a finding expert advice mode, and a product purchasing mode (column 5, lines 25-38).

As per claims 6 and 37, Breese discloses crawling network documents, wherein the crawling comprises parsing crawled documents for links, calculating probable user interest in the parsed links using the learning machine, and preferentially following links likely to be of interest to the user (column 9, lines 51-67, column 10, lines 1-27, 38-55).

As per claims 7 and 38, Breese discloses wherein the identified properties of the document  $d$  comprise a user  $u$ -independent property selected from the group consisting of:

- A probability  $P(t|d)$  that the document  $d$  is of interest to users interested in a topic  $t$  (column 6, lines 38-45);
- A topic classifier discrete probability distribution  $P(t/d)$  (column 6, lines 38-45);

Art Unit: 2157

- A product model discrete probability distribution  $P(p/d)$  (column 6, lines 38-45);
- Product feature values extracted from the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- An author of the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- An age of the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- A list of documents linked to the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- A language of the document  $d$  (column 9, lines 50-67, column 10, lines 1-20);
- A number of users who have accessed the document  $d$  (column 11, lines 1-30);
- A number of users who have saved the document  $d$  in a favorite document list (column 11, lines 1-30);
- A list of users previously interested in the document  $d$  (column 11, lines 1-30).

As per claims 8 and 39, Breese discloses wherein the parameters of the learning machine define a user  $u$ -dependent function selected from the group consisting of:

- A user topic probability distribution  $P(t/u)$  representing interests of the user  $u$  in various topics  $t$  (column 6, lines 38-45);
- A user product probability distribution  $P(p/u)$  representing interests of the user  $u$  in various products  $p$  (column 6, lines 38-45);
- A user product feature probability distribution  $P(F/u, p)$  representing interests of the user  $u$  in various features  $f$  of each of the various products  $p$  (column 6, lines 38-45, column 9, lines 50-67, column 10, lines 1-20);

Art Unit: 2157

- A website probability distribution  $P(s/u)$  representing interests of the user  $u$  in various websites  $s$  (column 8, lines 1-27);
- A cluster probability distribution  $P(c(u)/u)$  representing similarity of the user  $u$  to users in various clusters  $c(u)$  (column 9, lines 25-40);
- A phrase model probability distribution  $P(w/u)$  representing interests of the user  $u$  in various phrases  $w$  (column 9, lines 25-40);
- An information theory based measure  $I(lw; lu)$  representing mutual information between various phrases  $w$  and the user  $u$  (column 9, lines 25-40);
- An information theory based measure  $I(lt; lu)$  representing mutual information between various topics  $t$  and the user  $u$  (column 6, lines 38-45);
- An information theory based measure  $I(ls; lu)$  representing mutual information between various websites  $s$  and the user  $u$  (column 8, lines 1-27);
- An information theory based measure  $I(lp; lu)$  representing mutual information between various products  $p$  and the user  $u$  (column 8, lines 1-27);
- An information theory based measure  $I(lf; lu)$  representing mutual information between various features  $f$  of each of the various products  $p$  and the user  $u$  (column 8, lines 1-27).

As per claims 9 and 40, Breese discloses wherein the parameters of the learning machine define:

- A user product probability distribution  $P(p/u)$  representing interests of the user  $u$  in various products  $p$  (column 8, lines 1-27);

Art Unit: 2157

- A user product feature probability distribution  $P(f/u, p)$  representing interests of the user  $u$  in various features  $f$  of each of the various products  $p$  (column 8, lines 1-27);
- Estimating a probability  $P(u/d, \text{product described}=p)$  that a document  $d$  that describes a product  $p$  is of interest to the user  $u$ , wherein the probability is estimated in part from the user product probability distribution and the user product feature probability distribution (column 8, lines 1-27).

As per claims 10 and 41, Breese discloses recommending products to the user based on the probability  $P(u/d, \text{product described}=p)$  (column 6, lines 38-45).

As per claims 11 and 42, Breese discloses estimating a posterior probability  $P(u/d, q)$  that the document  $d$  is of interest to the user  $u$ , given a query  $q$  submitted by the user (column 8, lines 44-67).

As per claims 12 and 43, Breese discloses wherein estimating the posterior probability comprises estimating a probability  $P(q/d, u)$  that the query  $q$  is expressed by the user  $u$  with an information need in the document  $d$  (column 8, lines 44-67).

As per claims 13 and 44, Breese discloses applying the identified properties of the document  $d$  to a learning machine having product parameters characterizing a product  $p$  to estimate a probability  $P(p/d)$  that the document  $d$  refers to the product  $p$  (column 6, lines 38-45).

Art Unit: 2157

As per claims 14 and 45, Breese discloses updating the product parameters based on the identified properties of the document  $d$  and the estimated probability  $P(p/d)$  (column 8, lines 33-36, 44-46).

As per claims 15 and 46, Breese discloses initializing the product parameters based on a set of documents associated with the product  $p$  (column 8, lines 15-50).

As per claims 16 and 47, Breese discloses clustering multiple users into clusters of similar users, wherein the clustering comprises calculating distances between User Models, and selecting similar users based on the calculated distances between User Models (column 9, lines 25-40).

As per claims 17 and 48, discloses calculating relative entropy values between User Models of multiple users, and clustering together users based on the calculated relative entropy values (column 9, lines 25-40).

As per claims 18 and 49, Breese discloses wherein the parameters defining the User Model comprise calculated distances between the User Model and User Models of users similar to the user (column 11, columns 13-14).

Art Unit: 2157

As per claim 19 and 50, Breese discloses selecting in a group of users an expert user in an area expertise, wherein selecting the expert user comprises finding an expert User Model among User Models of the group of users, such that the expert User Model indicates a strong interest of the expert user in a document associated with the area of expertise (column 9, lines 20-50).

As per claim 20 and 51, Breese discloses parsing the document  $d$  for hyperlinks, and separately estimating for each of the hyperlinks a probability that the hyperlink is of interest to the user  $u$  (column 9, lines 51-67, column 10, lines 1-27, 38-55).

As per claims 21 and 52, Breese discloses sending to a third party web server user interest information derived from the User Model, whereby the third party web server may customize its interaction with the user (column 17, lines 6-25).

As per claims 22 and 53, Breese discloses wherein the monitored user interactions include a sequence of interaction times (column 9, lines 63-67).

As per claims 23 and 54, Breese discloses initializing the User Model using information selected from the group consisting of a set of documents provided by the user, a web browser history file associated with the user, a web browser bookmarks file associated with the user, ratings by the user of a set of documents, and previous product purchases made by the user (column 5, lines 25-38).

Art Unit: 2157

As per claims 24 and 55, Breese discloses modifying the User Model based on User Model modification requests provided by the user (column 8, lines 15-60).

As per claims 25 and 56, Breese discloses providing to the user a score for a document identified by the user, wherein the score is derived from the estimated probability (column 12, lines 49-55, column 13, lines 1-5).

As per claims 26 and 57, Breese discloses providing to the user a 3D map of a hyperlinked document collection, wherein the 3D map indicates a user interest in each document (column 5, lines 25-38).

As per claims 27 and 58, Breese discloses temporarily using a User Model that is built from a set of predetermined parameters of a profile selected by the user (column 8, lines 15-60).

As per claims 28 and 59, Breese discloses initializing the User Model by selecting a set of predetermined parameters of a prototype user selected by the user (column 8, lines 15-60).



As per claims 29 and 60, Breese discloses updating the predetermined parameters of the prototype user based on actions of users similar to the prototype user (column 8, lines 33-36, 44-46).

As per claims 30 and 61, Breese discloses identifying a set of users interest in the document d (column 16, lines 34-42).

As per claims 31 and 62, Breese discloses calculating a range of interests in the document d for the identified set of users (column 16, lines 34-42).

### ***Conclusion***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Barbara N Burgess whose telephone number is (703) 305-3366. The examiner can normally be reached on M-F (8:00am-4:00pm).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Ettinene can be reached on (703) 308-7562. The fax phone numbers for the organization where this application or proceeding is assigned are (703) 746-7239 for regular communications and (703) 746-7240 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-3900.

Barbara N Burgess  
Examiner  
Art Unit 2157

\*\*\*

June 2, 2003



**SALEH NAJJAR  
PRIMARY EXAMINER**

**Notice of References Cited**

Application/Control No.  
09/597,975

Applicant(s)/Patent Under  
Reexamination  
KONIG ET AL.

Examiner  
Barbara N Burgess

Art Unit  
2157

Page 1 of 1

**U.S. PATENT DOCUMENTS**

*	Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
A	US-6,006,218	12-1999	Breese et al.	707/3
B	US-			
C	US-			
D	US-			
E	US-			
F	US-			
G	US-			
H	US-			
I	US-			
J	US-			
K	US-			
L	US-			
M	US-			

**FOREIGN PATENT DOCUMENTS**

*	Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
N					
O					
P					
Q					
R					
S					
T					

**NON-PATENT DOCUMENTS**

*	Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
U	
V	
W	
X	

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)  
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

FORM PTO-1449 U.S. DEPARTMENT OF COMMERCE	ATTY. DOCKET NO. UTO-101	SERIAL NO.
<b>LIST OF PRIOR ART CITED BY APPLICANT</b> (Use several sheets if necessary)	APPLICANT KONIG, Yochai	
	FILING DATE	GROUP

**U.S. PATENT DOCUMENTS**

EXAMINER INITIAL	A	B	C	D	E	F	G	H	I	J	DOCUMENT NUMBER	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
BB											5 9 3 3 8 2 7	8/3/99	Cole et al.	707	10	
BB											5 9 9 9 9 7 5	12/7/99	Kittaka et al.	709	224	
BB											6 0 0 6 2 1 8	12/21/99	Breese et al.	707	3	
BB											5 7 5 4 9 3 9	5/19/98	Herz et al.	455	4.2	
BB											6 0 4 1 3 1 1	3/21/00	Chislenko et al.	705	27	
BB											5 9 1 8 0 1 4	6/29/99	Robinson	395	200.49	
BB											5 7 0 4 0 1 7	12/30/97	Heckerman et al.	395	61	
BB											6 0 2 9 1 6 1	2/22/00	Lang et al.	707	1	
BB											5 8 6 7 7 9 9	2/2/99	Lang et al.	707	1	
BB											5 9 8 3 2 1 4	11/9/99	Lang et al.	707	1	

**FOREIGN PATENT DOCUMENTS**

EXAMINER INITIAL	A	B	C	D	E	F	G	H	I	J	DOCUMENT NUMBER	DATE	COUNTRY	CLASS	SUBCLASS	TRANSLATION		
																YES	NO	

**OTHER PRIOR ART (Including Author, Title, Date, Pertinent Pages, Etc.)**

EXAMINER INITIAL	K	Mobasher, B., <i>Automatic personalization based on web usage mining</i> , <a href="http://maya.cs.depaul.edu/~mobasher/personalization">http://maya.cs.depaul.edu/~mobasher/personalization</a> .
	L	
	M	
	N	

EXAMINER <i>Ralph Byer</i>	DATE CONSIDERED <i>3-15-03</i>
-------------------------------	-----------------------------------

\*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

L Number	Hits	Search Text	DB	Time stamp
1	1	6006218.pn.	USPAT; US-PGPUB	2003/06/02 08:57
-	1	5933827.pn.	USPAT; US-PGPUB	2003/06/02 08:57
-	0	599997.pn.	USPAT; US-PGPUB	2003/05/19 15:48
-	1	5999975.pn.	USPAT; US-PGPUB	2003/05/19 15:49
-	1	6006218.pn.	USPAT; US-PGPUB	2003/05/19 15:50
-	1	5754939.pn.	USPAT; US-PGPUB	2003/05/19 15:52
-	1	6041311.pn.	USPAT; US-PGPUB	2003/05/19 15:53
-	1	5918014.pn.	USPAT; US-PGPUB	2003/05/19 15:55
-	1	5704017.pn.	USPAT; US-PGPUB	2003/05/19 15:55
-	1	6029161.pn.	USPAT; US-PGPUB	2003/05/19 15:56
-	1	5867799.pn.	USPAT; US-PGPUB	2003/05/19 15:58
-	1	5983214.pn.	USPAT; US-PGPUB	2003/05/30 15:18
-	2152	709/224.ccls.	USPAT; US-PGPUB	2003/05/30 15:18
-	1359	709/224.ccls. and (@ad<19991228 or @prad<19991228)	USPAT; US-PGPUB	2003/05/30 15:19
-	273	709/224.ccls. and (@ad<19991228 or @prad<19991228) and monitor\$5 same (user\$1 or customer\$1 or computer\$1 or device\$1 or workstation\$1) same (interaction\$1 or key\$5 or mouse or click\$4)	USPAT; US-PGPUB	2003/05/30 15:23
-	4	709/224.ccls. and (@ad<19991228 or @prad<19991228) and monitor\$5 same (user\$1 or customer\$1 or computer\$1 or device\$1 or workstation\$1) same (interaction\$1 or key\$5 or mouse or click\$4) and (database or file\$1 or profile\$1) and (advertis\$5 or email\$1 or banner\$1 or coupon\$1 or incentive\$1) same (probabilit\$3 or chance or guess)	USPAT; US-PGPUB	2003/05/30 15:27
-	11	709/224.ccls. and (@ad<19991228 or @prad<19991228) and monitor\$5 same (user\$1 or customer\$1 or computer\$1 or device\$1 or workstation\$1) same (interaction\$1 or key\$5 or mouse or click\$4) and (database or file\$1 or profile\$1) and (advertis\$5 or email\$1 or banner\$1 or coupon\$1 or incentive\$1 or distribut\$4) same (probabilit\$3 or chance or guess)	USPAT; US-PGPUB	2003/05/30 15:28
-	13	709/224.ccls. and (@ad<19991228 or @prad<19991228) and monitor\$5 same (user\$1 or customer\$1 or computer\$1 or device\$1 or workstation\$1) same (interaction\$1 or key\$5 or mouse or click\$4) and (database or file\$1 or profile\$1) and (advertis\$5 or email\$1 or banner\$1 or coupon\$1 or incentive\$1 or distribut\$4) same (probabilit\$3 or chance or estimat\$4 or guess)	USPAT; US-PGPUB	2003/05/30 15:28

-	781	(@ad<19991228 or @prad<19991228) and monitor\$5 same (user\$1 or customer\$1 or computer\$1 or device\$1 or workstation\$1) same (interaction\$1 or key\$5 or mouse or click\$4) and (database or file\$1 or profile\$1) and (advertis\$5 or email\$1 or banner\$1 or coupon\$1 or incentive\$1 or distribut\$4) same (probabilit\$3 or chance or estimat\$4 or guess)	USPAT; US-PGPUB	2003/05/30 15:30
-	121	(@ad<19991228 or @prad<19991228) and monitor\$5 same (user\$1 or customer\$1 or computer\$1 or device\$1 or workstation\$1) same (interaction\$1 or key\$5 or mouse or click\$4) same (database or file\$1 or profile\$1) and (advertis\$5 or email\$1 or banner\$1 or coupon\$1 or incentive\$1 or distribut\$4) same (probabilit\$3 or chance or estimat\$4)	USPAT; US-PGPUB	2003/05/30 15:31



UNITED STATES DEPARTMENT OF COMMERCE  
Patent and Trademark Office

ASSISTANT SECRETARY AND COMMISSIONER  
OF PATENTS AND TRADEMARKS  
Washington, D.C. 20231

#5

CHANGE OF ADDRESS/POWER OF ATTORNEY

FILE LOCATION 21C1 SERIAL NUMBER 09597975 PATENT NUMBER

THE CORRESPONDENCE ADDRESS HAS BEEN CHANGED TO CUSTOMER # 30869

THE PRACTITIONERS OF RECORD HAVE BEEN CHANGED TO CUSTOMER # 30869

THE FEE ADDRESS HAS BEEN CHANGED TO CUSTOMER # 30869

ON 05/15/03 THE ADDRESS OF RECORD FOR CUSTOMER NUMBER 30869 IS:

LUMEN INTELLECTUAL PROPERTY SERVICES  
45 CABOT AVENUE, SUITE 110  
SANTA CLARA CA 95051

AND THE PRACTITIONERS OF RECORD FOR CUSTOMER NUMBER 30869 ARE:

39299 39894 50000 50142

RECEIVED

MAY 21 2003

Technology Center 2100

PTO INSTRUCTIONS: PLEASE TAKE THE FOLLOWING ACTION WHEN THE  
CORRESPONDENCE ADDRESS HAS BEEN CHANGED TO CUSTOMER NUMBER:  
RECORD, ON THE NEXT AVAILABLE CONTENTS LINE OF THE FILE JACKET,  
'ADDRESS CHANGE TO CUSTOMER NUMBER'. LINE THROUGH THE OLD  
ADDRESS ON THE FILE JACKET LABEL AND ENTER ONLY THE 'CUSTOMER  
NUMBER' AS THE NEW ADDRESS. FILE THIS LETTER IN THE FILE JACKET.  
WHEN ABOVE CHANGES ARE ONLY TO FEE ADDRESS AND/OR PRACTITIONERS  
OF RECORD, FILE LETTER IN THE FILE JACKET.  
THIS FILE IS ASSIGNED TO GAU 2157.





**BEST AVAILABLE COPY**

<b>SEARCHED</b>			
Class	Sub.	Date	Exmr.
709	224	5-30-03	GB
/			

<b>INTERFERENCE SEARCHED</b>			
Class	Sub.	Date	Exmr.

<b>SEARCH NOTES (INCLUDING SEARCH STRATEGY)</b>		
	Date	Exmr.
EAST Text Search	5-30-03	GB
EAST Image Search	5-30-03	
EAST Text Search	1-15-04 1-16-04	GB GB
<u>Saleh Najjar</u> - discussed day 2 103 w/ 2 reference (Brose)	1-14-04	GB
<u>Saleh Najjar</u> - go find - 103 w/ 2 reference	5-19-04	GB

(RIGHT OUTSIDE)

POSITION	INITIALS	ID NO.	DATE
FEE DETERMINATION	W	70530	06-29-00
O.I.P.E. CLASSIFIER		10	7-6-00
FORMALITY REVIEW	AL		8/16/00
RESPONSE FORMALITY REVIEW			

**INDEX OF CLAIMS**

- ✓ ..... Rejected
- = ..... Allowed
- (Through numeral)... Canceled
- + ..... Restricted
- N ..... Non-elected
- I ..... Interference
- A ..... Appeal
- O ..... Objected

Claim	Final	Original	Date
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			

Claim	Final	Original	Date
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

Claim	Final	Original	Date
101			
102			
103			
104			
105			
106			
107			
108			
109			
110			
111			
112			
113			
114			
115			
116			
117			
118			
119			
120			
121			
122			
123			
124			
125			
126			
127			
128			
129			
130			
131			
132			
133			
134			
135			
136			
137			
138			
139			
140			
141			
142			
143			
144			
145			
146			
147			
148			
149			
150			

BEST AVAILABLE COPY

If more than 150 claims or 10 actions  
staple additional sheet here

(LEFT INSIDE)

06/20/00

1c813 U.S. PTO

A

<b>UTILITY PATENT APPLICATION TRANSMITTAL</b>	Attorney Docket No. <b>UTO-101</b>	Total Pages
	First Named Inventor <b>YOCHAI KONIG</b>	
	Title <b>AUTOMATIC, PERSONALIZED ONLINE INFORMATION AND PRODUCT SERVICES</b>	

1c813 U.S. PTO  
06/20/00

APPLICATION ELEMENTS	ACCOMPANYING APPLICATION PARTS
1. <input checked="" type="checkbox"/> Fee Transmittal Form	8. <input checked="" type="checkbox"/> Assignment cover sheet and document(s)
2. <input checked="" type="checkbox"/> Specification Total Pages: [57]	9. <input checked="" type="checkbox"/> Power of Attorney by Assignee [ ] with CFR 3.73(b) statement
3. <input checked="" type="checkbox"/> Drawing(s) Total Sheets: [19]	10. [ ] English Translation Document
4. <input checked="" type="checkbox"/> Oath or Declaration Total Pages: [ ]	11. <input checked="" type="checkbox"/> IDS/PTO-1449 [X] with copies of cited references
a. <input checked="" type="checkbox"/> Newly executed (original or copy)	12. [ ] Preliminary Amendment
b. [ ] Copy from a prior application 1.63(d) (complete Box 17 and note Box 5 below)	13. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503)
i. [ ] Signed statement deleting inventors named in the prior application, see CFR 1.63(d)(2) and 1.33(b)	14. <input checked="" type="checkbox"/> Small Entity Statement [ ] Statement filed in prior application Status still proper and desired
5. [ ] Incorporation by Reference (if 4b is checked) The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated therein by reference.	15. [ ] Certified Copy of Priority Document(s) (if foreign priority is claimed)
6. [ ] Microfiche Computer Program (Appendix)	16. [ ] Other:
7. [ ] Nucleotide/Amino Acid Sequence Submission (all the following are necessary)	
a. [ ] Computer Readable Copy	
b. [ ] Paper Copy (identical to computer copy)	
c. [ ] Statement verifying identity of above copies	


17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:  
 continuation     divisional     CIP    ...of prior application No.: [ ]

18. CORRESPONDENCE ADDRESS					
NAME	<b>RENA KAMINSKY LUMEN INTELLECTUAL PROPERTY SERVICES</b>				
ADDRESS	<b>45 CABOT AVENUE, SUITE 110</b>				
CITY	<b>SANTA CLARA</b>	STATE	<b>CA</b>	ZIP CODE	<b>95051</b>
COUNTRY	<b>USA</b>	TELEPHONE	<b>(408) 260-7300</b>	FAX	<b>(408) 260-7301</b>

**Certificate of Mailing by "Express Mail"**

I hereby certify that I am mailing this correspondence on the date indicated below to the ASSISTANT COMMISSIONER FOR PATENTS, BOX PATENT APPLICATION, WASHINGTON, DC 20231 using the "Express Mail Post Office to Addressee" service of the United States Postal Service under 37 CFR 1.10.

DATE OF MAILING: June 20, 2000

  
**RENA KAMINSKY**  
REG. NO. P-46,818

**EXPRESS MAIL LABEL NC EK742114277US**

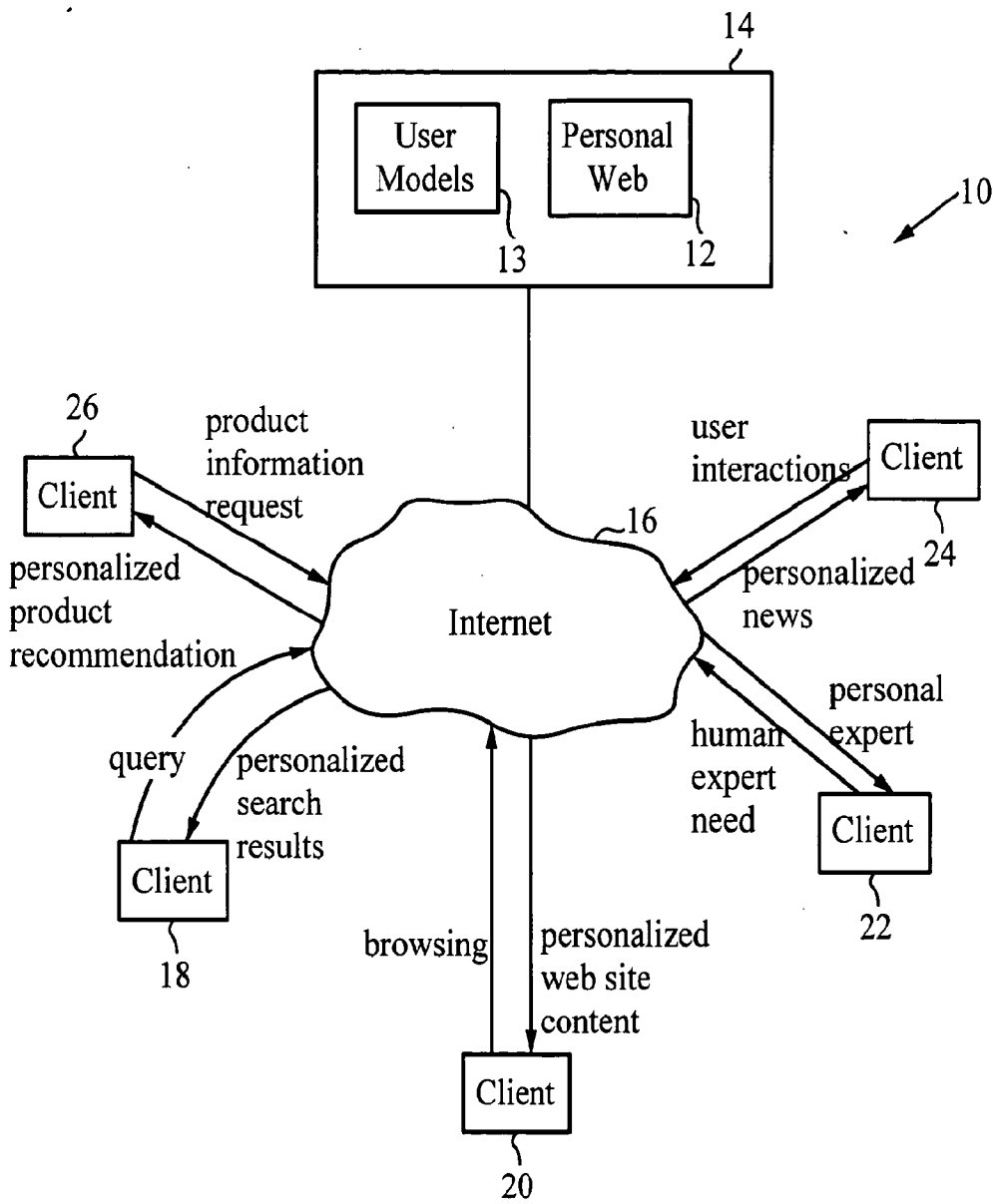
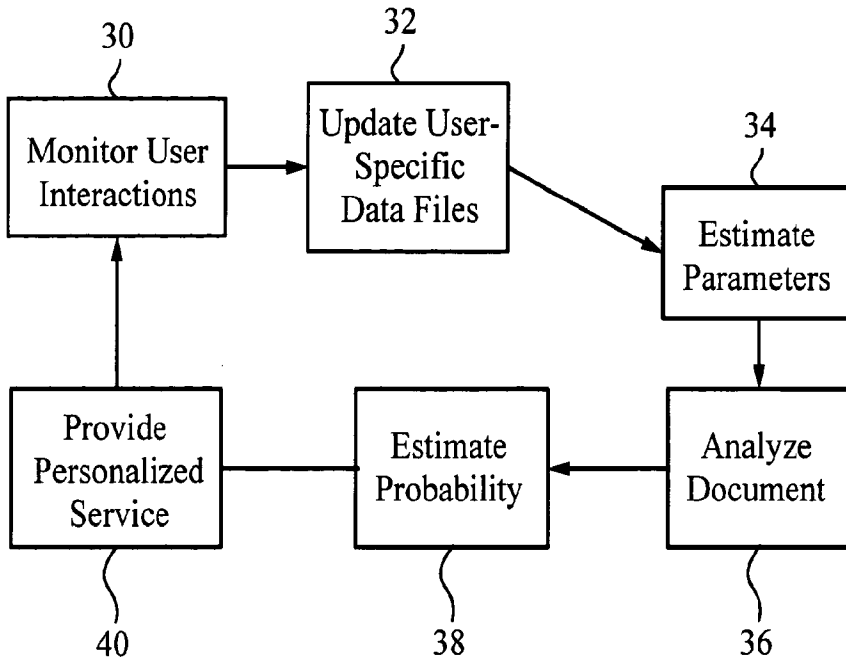
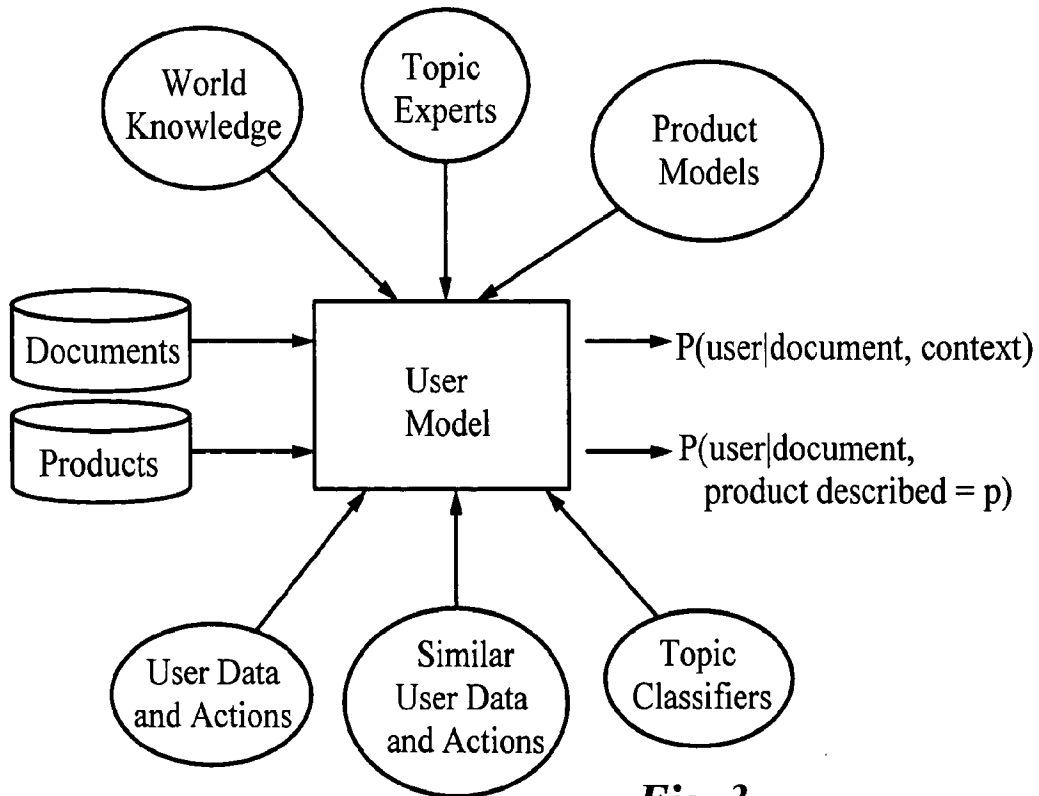


Fig. 1



*Fig. 2*



*Fig. 3*

Informative Word/Phrase List

Word ID	Word Grade	Last Access Time	Number of Accesses
Vegan	0.86	3/6/2000 12:22:41	173
Parasail	0.72	4/15/2000 18:51:27	220

*Fig. 4A*

Web Site Distribution

Site ID	Site Probability	Last Access Time	Number of Accesses
herring.com	0.61	5/1/2000 19:15:21	152
Java.com	0.43	4/24/2000 3:16:18	460

*Fig. 4B*

User Topic Distribution

Topic ID	Topic Parent	Topic Probability	Last Access Time	Number of Accesses
Computers	Industries	0.6	12/2/1999 1:21:22	74
Publishing	Industries	0.31	1/2/2000 6:25:31	62

*Fig. 4C*

User Product Distribution

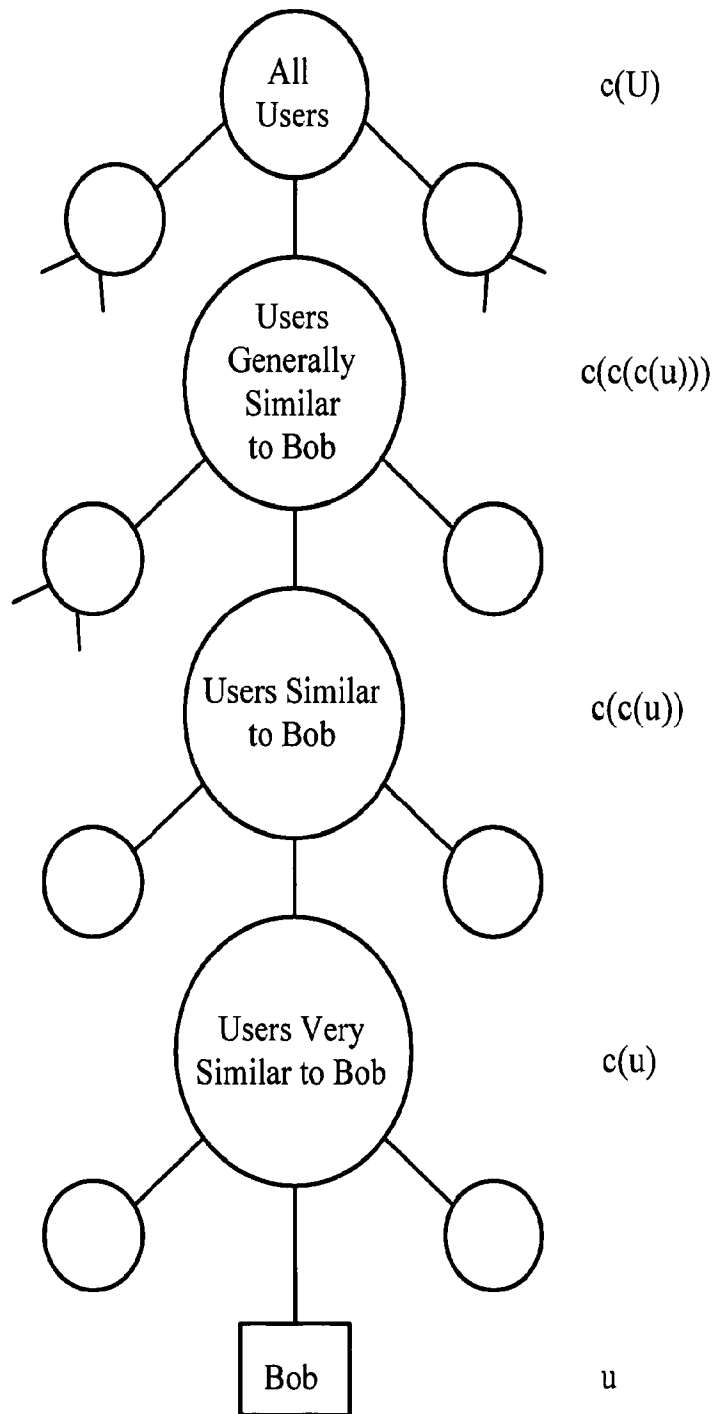
Product ID	Product Parent	Product Probability	Last Purchase Time	Number of Purchases	Last Access Time	Number of Accesses
3Com Palm 3E	Without Keyboards	0.73	12/16/1999 17:21:21	1	5/2/2000 16:01:21	78
Without Keyboards	Handhelds/PDAs	0.81	12/16/1999 17:21:21	1	3/15/2000 17:21:21	98

*Fig. 4D*

User Product Feature Distribution

Product ID	Feature ID	Value	Value Probability
Webcams	Interface	PC Card	0.7
Webcams	Interface	Serial	0.2

*Fig. 4E*



**Fig. 5A**

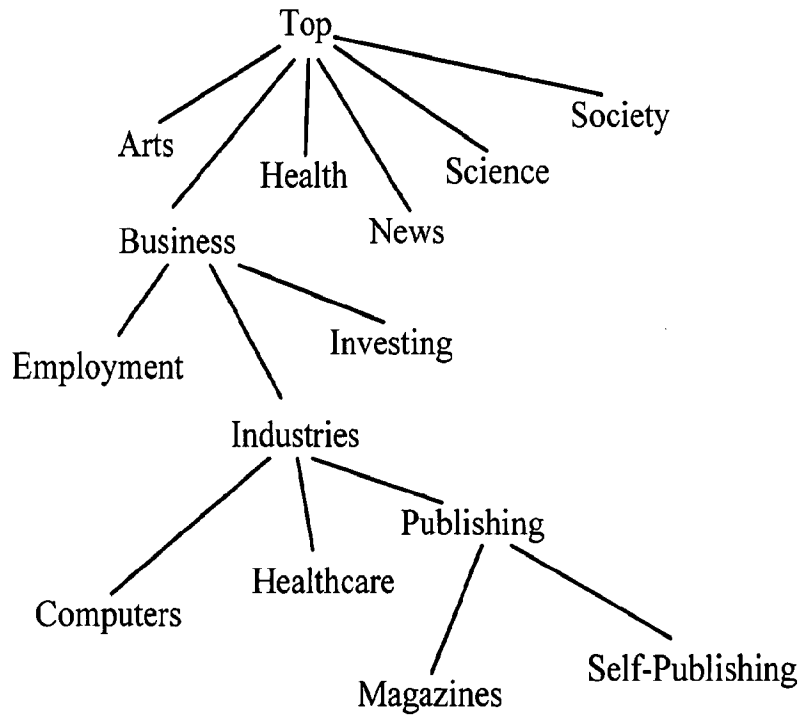




User Fuzzy Cluster Tree

Cluster ID	Cluster Parent ID	Cluster Probability
Bob	C1	0.3
Bob	C2	0.2
Bob	C3	0.1
Bob	C4	0.4
C1	C11	0.2

*Fig. 6B*



*Fig. 7*







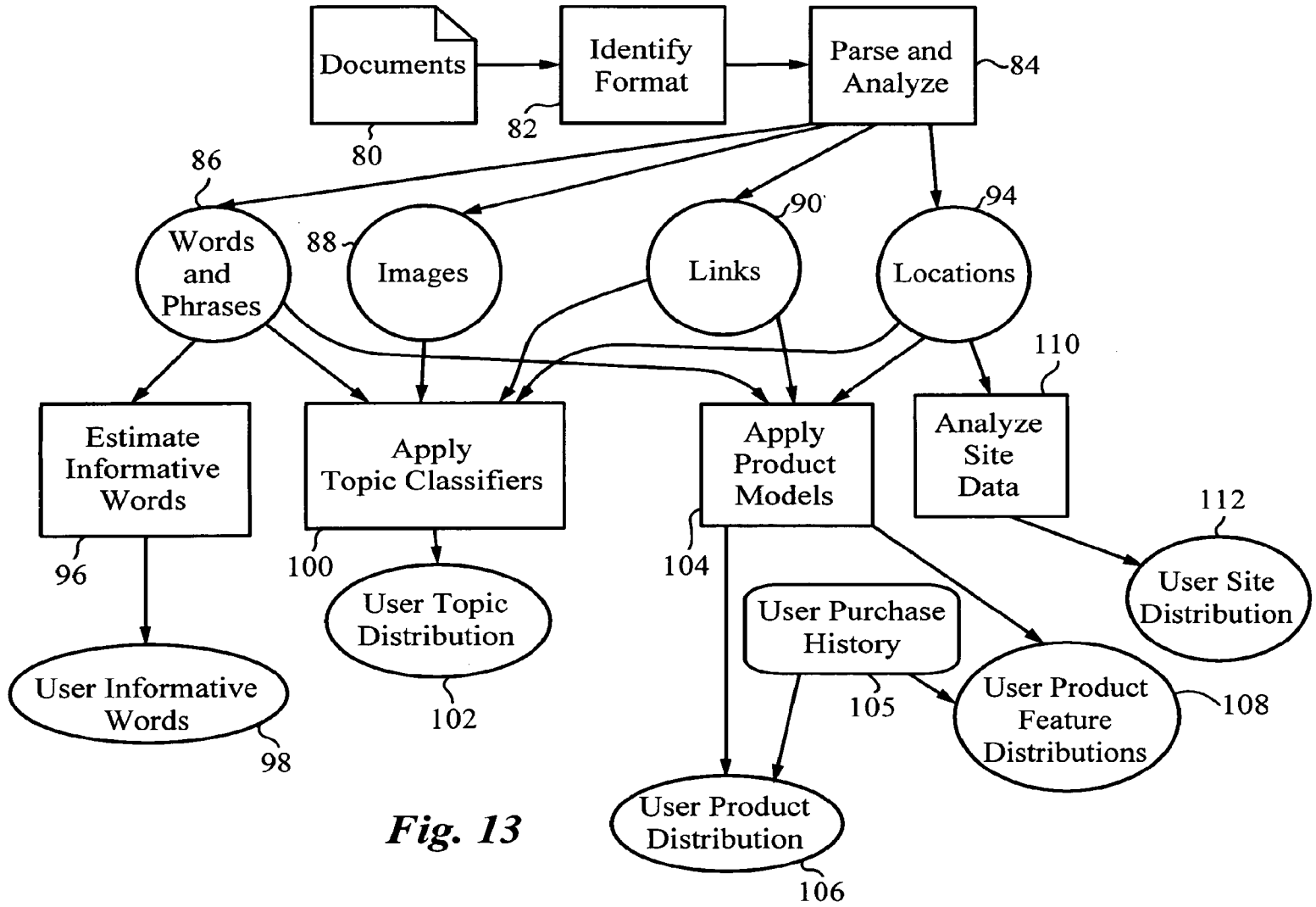


Fig. 13

User Recently Accessed Butter

Document ID	Access Time	Interaction Type	Context	Degree of Interest
www.herring.com/insider	5/12/2000 14:37:21	Navigation	bookmark access	positive, followed 3 links 12 minutes
www.m-w.com	5/12/2000 15:08:21	Search	query "dictionary"	positive, followed 5 links bookmarked, 21 minutes

*Fig. 14*

User Site Candidate Table

Site Name	Number of Access	Last Access Time
www.herring.com	157	5/12/2000 14:37:21
www.m-w.com	162	5/12/2000 15:08:21

*Fig. 15A*

User Word Candidate Table

Word ID	Word Spelling	Word Spelling	Word Grade	Last Access Time
Cytochrome	Cytochrome	Cytocrome	0.67	4/16/2000 7:10:01
Hyperbilirubinemia	Hyperbilirubinemia	Hyperbilirubenema	0.58	4/27/2000 12:18:42

*Fig. 15B*

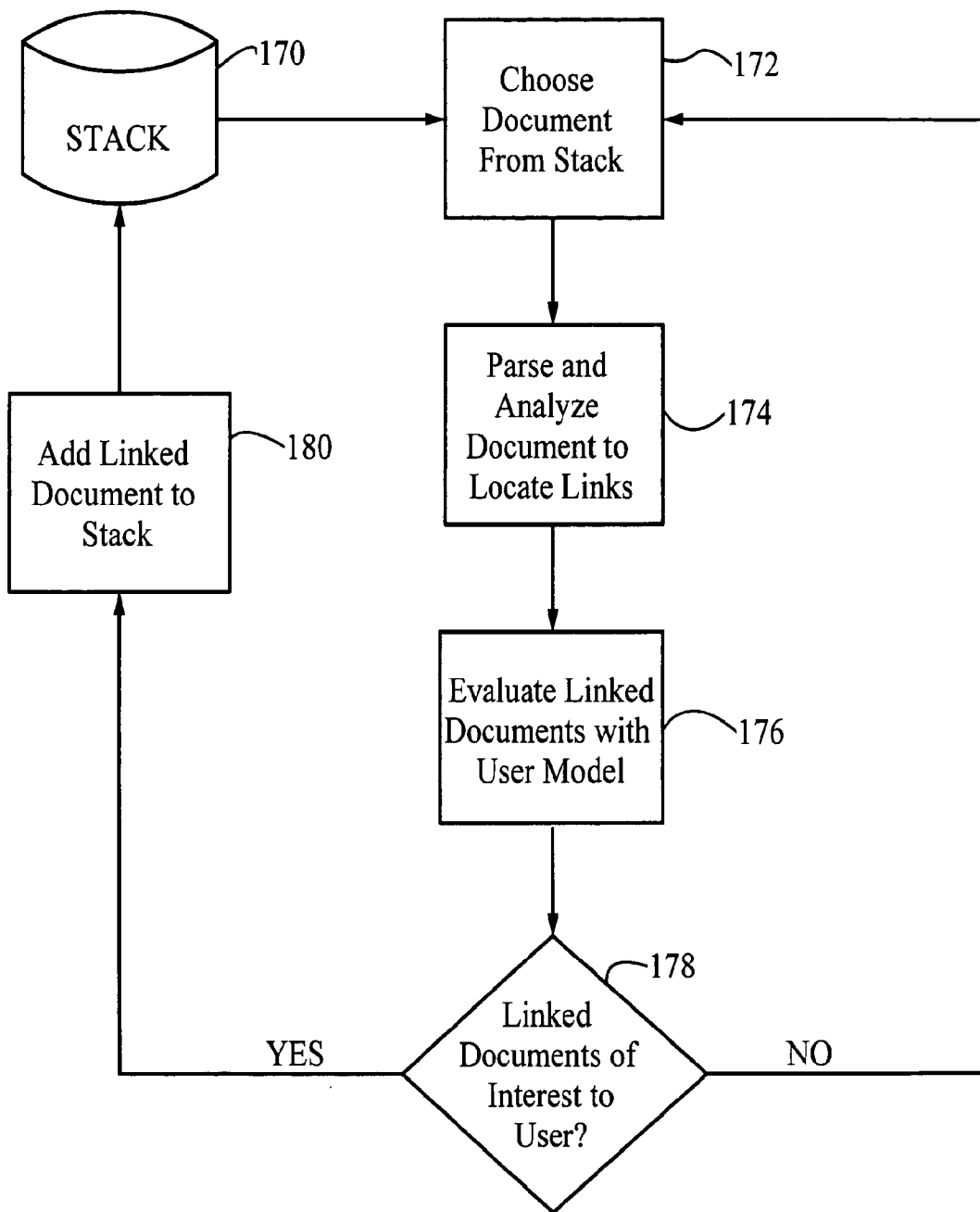
User Recently Purchased Products

Product ID	Parent Node	Purchase Time	Purchase Source
Panasonic SL-502	Discmans	5/1/2000 16:01:04	ebyweb.com
Hitachi VM6500A	Camcorders	5/3/2000 18:19:21	supremevideo.com

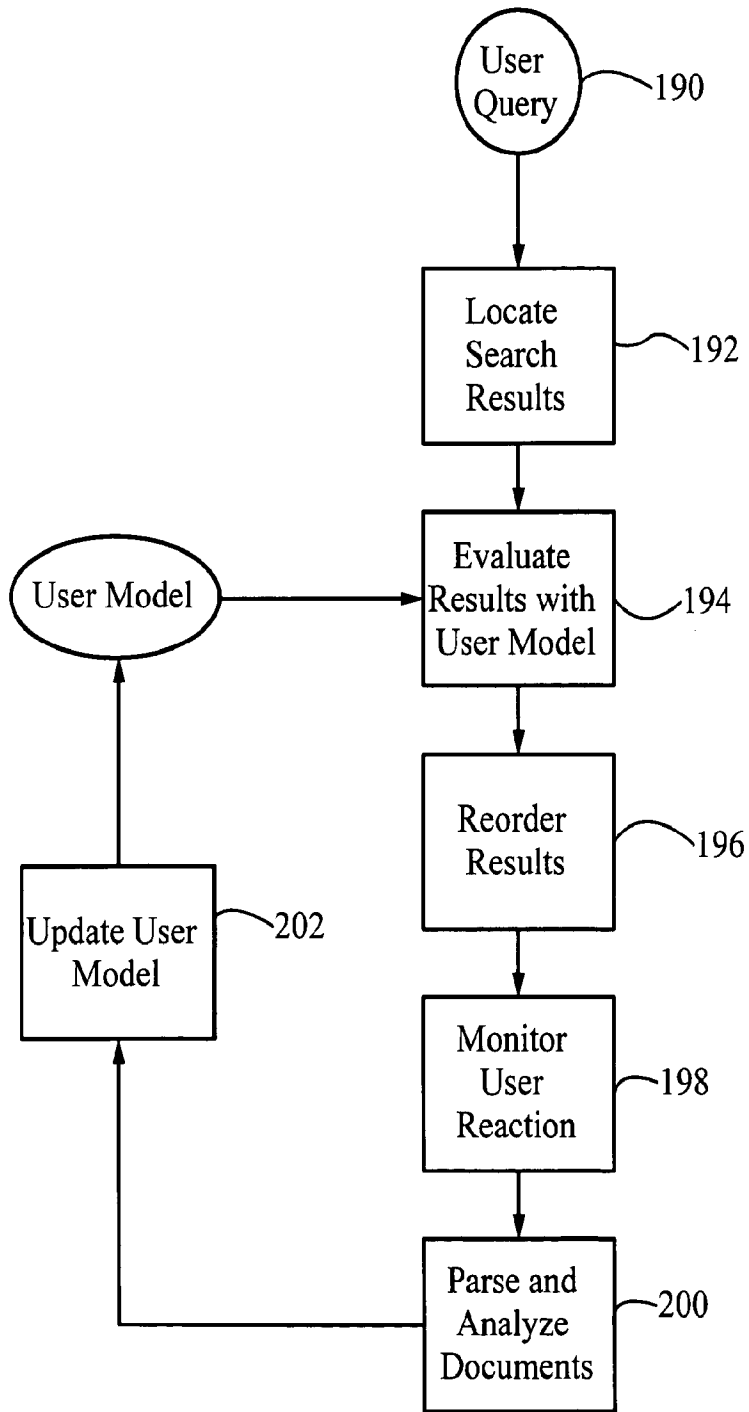
*Fig. 16*





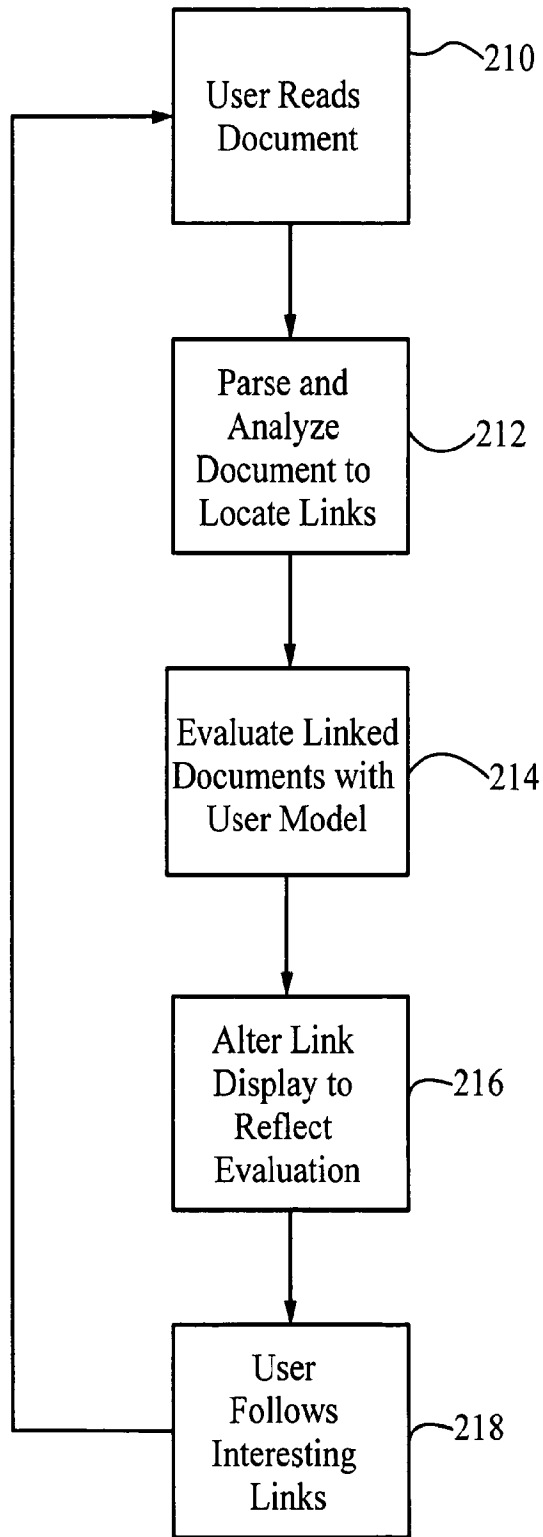


**Fig. 18**

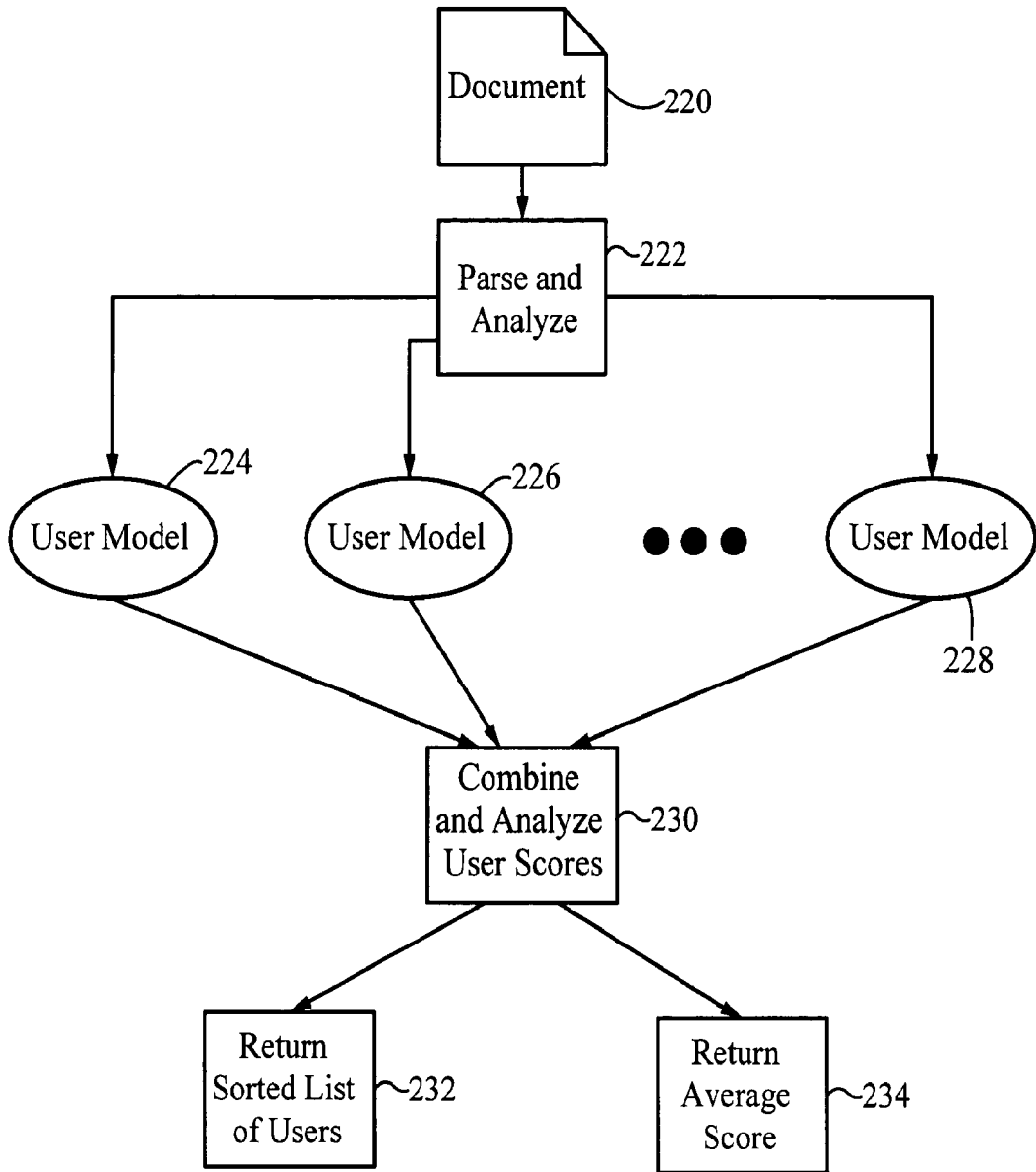


*Fig. 19*

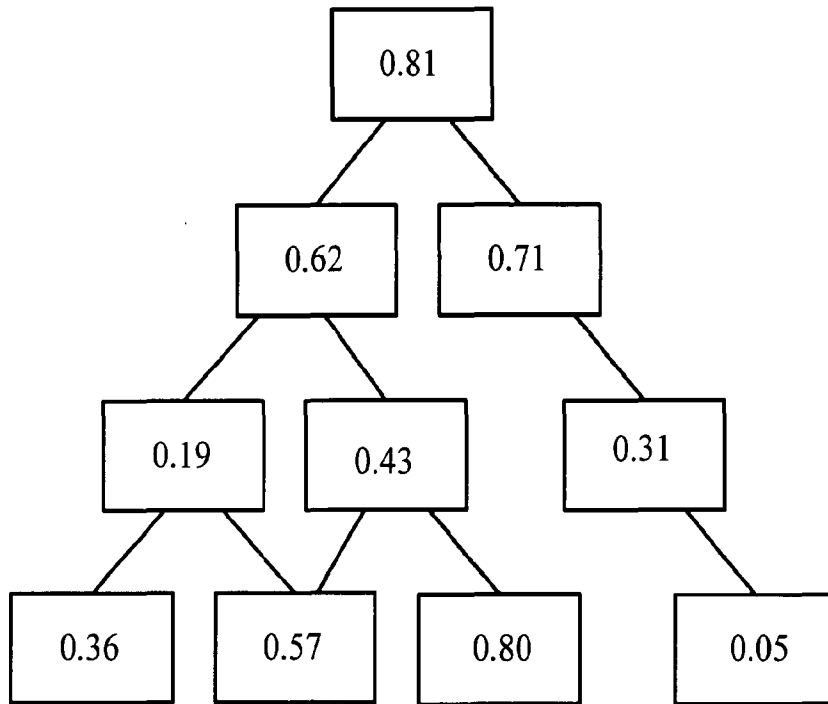
17/19



*Fig. 20*



*Fig. 21*



*Fig. 22*

0.81  
0.62  
0.71  
0.19  
0.43  
0.31  
0.36  
0.57  
0.80  
0.05

Patent Application of  
**Yochai Konig, Roy Twersky, and Michael R. Berthold**

for

**Automatic, Personalized Online Information and Product Services**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims the benefit of U.S. Provisional Application No. 60/173,392 filed 12/28/99, which is herein incorporated by reference.

**FIELD OF THE INVENTION**

This invention relates generally to methods for personalizing a user's interaction with information in a computer network. More particularly, it relates to methods for predicting user interest in documents and products using a learning machine that is continually updated based on actions of the user and similar users.

**BACKGROUND ART**

The amount of static and dynamic information available today on the Internet is staggering, and continues to grow exponentially. Users searching for information, news, or products and services are quickly overwhelmed by the volume of information, much of it useless and uninformative. A variety of techniques have been developed to organize, filter, and search for information of interest to a particular user. Broadly, these methods can be divided into information filtering techniques and collaborative filtering techniques.

Information filtering techniques focus on the analysis of item content and the development of a personal user interest profile. In the simplest case, a user is characterized by a set of documents, actions regarding previous documents, and user-defined parameters, and new documents are characterized and compared with the user profile. For example, U.S. Patent No. 5,933,827, issued to Cole et al., discloses a system for identifying new web pages of interest to a user. The user is characterized simply by a set of categories, and new documents are categorized and compared with the user's profile. U.S. Patent No. 5,999,975, issued to Kittaka et al., describes an online information providing scheme that characterizes users and documents by a set of attributes, which are compared and updated base on user selection of particular

Sub  
AI

documents. U.S. Patent No. 6,006,218, issued to Breese et al., discloses a method for retrieving information based on a user's knowledge, in which the probability that a user already knows of a document is calculated based on user-selected parameters or popularity of the document. U.S. Patent No. 5,754,939, issued to Herz et al., discloses a method for identifying objects of interest to a user based on stored user profiles and target object profiles. Other techniques rate documents using the TFIDF (term frequency, inverse document frequency) measure. The user is represented as a vector of the most informative words in a set of user-associated documents. New documents are parsed to obtain a list of the most informative words, and this list is compared to the user's vector to determine the user's interest in the new document.

Existing information filtering techniques suffer from a number of drawbacks. Information retrieval is typically a two step process, collection followed by filtering; information filtering techniques personalize only the second part of the process. They assume that each user has a personal filter, and that every network document is presented to this filter. This assumption is simply impractical given the current size and growth of the Internet; the number of web documents is expected to reach several billion in the next few years. Furthermore, the dynamic nature of the documents, e.g., news sites that are continually updated, makes collection of documents to be filtered later a challenging task for any system. User representations are also relatively limited, for example, including only a list of informative words or products or user-chosen parameters, and use only a single mode of interaction to make decisions about different types of documents and interaction modes. In addition, information filtering techniques typically allow for extremely primitive updating of a user profile, if any at all, based on user feedback to recommended documents. As a user's interests change rapidly, most systems are incapable of providing sufficient personalization of a user's experience.

Collaborative filtering methods, in contrast, build databases of user opinions of available items, and then predict a user opinion based on the judgments of similar users. Predictions typically require offline data mining of very large databases to recover association rules and patterns; a significant amount of academic and industrial research is focussed on developing more efficient and accurate data mining techniques. The earliest collaborative filtering systems required explicit ratings by the users, but existing systems are implemented without the user's knowledge by observing user actions. Ratings are inferred from, for example, the amount of time a user spends reading a document or whether a user purchases a particular product. For example, an automatic personalization



method is disclosed in B. Mobasher et al., "Automatic Personalization Through Web Usage Mining," Technical Report TR99-010, Department of Computer Science, Depaul University, 1999. Log files of documents requested by users are analyzed to determine usage patterns, and online recommendations of pages to view are supplied to users based on the derived patterns and other pages viewed during the current session.

Recently, a significant number of web sites have begun implementing collaborative filtering techniques, primarily for increasing the number and size of customer purchases. For example, Amazon.com™ has a "Customers Who Bought" feature, which recommends books frequently purchased by customers who also purchased a selected book, or authors whose work is frequently purchased by customers who purchased works of a selected author. This feature uses a simple "shopping basket analysis"; items are considered to be related only if they appear together in a virtual shopping basket. Net Perceptions, an offshoot of the GroupLens project at the University of Minnesota, is a company that provides collaborative filtering to a growing number of web sites based on data mining of server logs and customer transactions, according to predefined customer and product clusters.

Numerous patents disclose improved collaborative filtering systems. A method for item recommendation based on automated collaborative filtering is disclosed in U.S. Patent No. 6,041,311, issued to Chislenko et al. Similarity factors are maintained for users and for items, allowing predictions based on opinions of other users. In an extension of standard collaborative filtering, item similarity factors allow predictions to be made for a particular item that has not yet been rated, but that is similar to an item that has been rated. A method for determining the best advertisements to show to users is disclosed in U.S. Patent No. 5,918,014, issued to Robinson. A user is shown a particular advertisement based on the response of a community of similar users to the particular advertisement. New ads are displayed randomly, and the community interest is recorded if enough users click on the ads. A collaborative filtering system using a belief network is disclosed in U.S. Patent No. 5,704,317, issued to Heckerman et al., and allows automatic clustering and use of non-numeric attribute values of items. A multi-level mindpool system for collaborative filtering is disclosed in U.S. Patent No. 6,029,161, issued to Lang et al. Hierarchies of users are generated containing clusters of users with similar properties.

Collaborative filtering methods also suffer from a number of drawbacks, chief of which is their inability to rate content of an item or incorporate user context. They are based only on user opinions; thus an item that has never been rated cannot be recommended or evaluated. Similarly, obscure items, which are rated by only a few users, are unlikely to be recommended. Furthermore, they require storage of a profile for every item, which is unfeasible when the items are web pages. New items cannot be automatically added into the database. Changing patterns and association rules are not incorporated in real time, since the data mining is performed offline. In addition, user clusters are also static and cannot easily be updated dynamically.

Combinations of information filtering and collaborative filtering techniques have the potential to supply the advantages provided by both methods. For example, U.S. Patent No. 5,867,799, issued to Lang et al., discloses an information filtering method that incorporates both content-based filtering and collaborative filtering. However, as with content-based methods, the method requires every document to be filtered as it arrives from the network, and also requires storage of a profile of each document. Both of these requirements are unfeasible for realistically large numbers of documents. An extension of this method, described in U.S. Patent No. 5,983,214, also to Lang et al., observes the actions of users on content profiles representing information entities. Incorporating collaborative information requires that other users have evaluated the exact content profile for which a rating is needed.

In summary, none of the existing prior art methods maintain an adaptive content-based model of a user that changes based on user behavior, allow for real-time updating of the model, operate during the collection stage of information retrieval, can make recommendations for items or documents that have never been evaluated, or model a user based on different modes of interaction.

### OBJECTS AND ADVANTAGES

Accordingly, it is a primary object of the present invention to provide a method of personalizing user interaction with network documents that maintains an adaptive content-based profile of the user.

It is another object of the invention to incorporate into the profile user behavior during different modes of interaction with information, thus allowing for cross-fertilization. Learning about the user interests in one mode benefits all other modes.

It is a further object of the invention to provide a method that jointly models the user's information needs and product needs to provide stronger performance in both modes.

It is an additional object of the invention to provide a method that personalizes both the collection and filtering stages of information retrieval to manage efficiently the enormous number of existing web documents.

It is another object of the invention to provide a method for predicting user interest in an item that incorporates the opinions of similar users without requiring storage and maintenance of an item profile.

It is a further object of the invention to provide an information personalization method that models the user as a function independent of any specific representation or data structure, and represents the user interest in a document or product independently of any specific user information need. This approach enables the addition of new knowledge sources into the user model.

It is an additional object of the present invention to provide a method based on Bayesian statistics that updates the user profile based on both negative and positive examples.

It is a further object of the invention to model products by analyzing all relevant knowledge sources, such as press releases, reviews, and articles, so that a product can be recommended even if it has never been purchased or evaluated previously.

## SUMMARY

These objects and advantages are attained by a computer-implemented method for providing automatic, personalized information services to a user. User interactions with a computer are transparently monitored while the user is engaged in normal use of the computer, and monitored interactions are used to update user-specific data files that include a set of documents associated with the user. Parameters of a learning machine, which define a User Model specific to the user, are estimated from the user-specific data files. Documents that are of interest and documents that are not of interest to the user are treated distinctly in estimating the parameters. The parameters are used to estimate a probability  $P(u|d)$  that a document is of interest to the user, and the estimated probability is then used to provide personalized information services to the user.

The probability is estimated by analyzing properties of the document and applying them to the learning machine. Documents of multiple distinct media types are analyzed, and identified properties include: the probability that the document is of interest to users who are interested in particular topics, a topic classifier probability distribution, a product model probability distribution, product feature values extracted from the document, the document author, the document age, a list of documents linked to the document, the document language, number of users who have accessed the document, number of users who have saved the document in a favorite document list, and a list of users previously interested in the document. All properties are independent of the particular user. The product model probability distribution, which indicates the probability that the document refers to particular products, is obtained by applying the document properties to a product model, a learning machine with product parameters characterizing particular products. These product parameters are themselves updated based on the document properties and on the product model probability distribution. Product parameters are initialized from a set of documents associated with each product.

User interactions are monitored during multiple distinct modes of user interaction with network data, including a network searching mode, network navigation mode, network browsing mode, email reading mode, email writing mode, document writing mode, viewing “pushed” information mode, finding expert advice mode, and product purchasing mode. Based on the monitored interactions, parameters of the learning machine are updated. Learning machine parameters define various user-dependent functions of the User Model, including a user topic probability distribution representing interests of the user in various topics, a user product probability distribution representing interests of the user in various products, a user product feature probability distribution representing interests of the user in various features of each of the various products, a web site probability distribution representing interests of the user in various web sites, a cluster probability distribution representing similarity of the user to users in various clusters, and a phrase model probability distribution representing interests of the user in various phrases. Some of the user-dependent functions can be represented as information theory based measures representing mutual information between the user and either phrases, topics, products, features, or web sites. The product and feature distributions can also be used to recommend products to the user.

The User Model is initialized from documents provided by the user, a web browser history file, a web browser bookmarks file, ratings by the user of a set of documents, or previous product purchases made by the user. Alternatively, the User Model may be initialized by selecting a set of predetermined parameters of a prototype user selected by the user. Parameters of the prototype user are updated based on actions of users similar to the prototype user. The User Model can be modified based on User Model modification requests provided by the user. In addition, the user can temporarily use a User Model that is built from a set of predetermined parameters of a profile selected by the user.

Distances between users are calculated to determine similar users, who are clustered into clusters of similar users. Parameters defining the User Model may include the calculated distances between the User Model and User Models of users within the user's cluster. Users may also be clustered based on calculated relative entropy values between User Models of multiple users.

A number of other probabilities can be calculated, such as a posterior probability  $P(u|d,q)$  that the document is of interest to the user, given a search query submitted by the user. Estimating the posterior probability includes estimating a probability that the query is expressed by the user with an information need contained in the document. In addition, the probability  $P(u|d,con)$  that the document is of interest to the user during a current interaction session can be calculated. To do so,  $P(u,con|d)/P(con|d)$  is calculated, where *con* represents a sequence of interactions during the current interaction session or media content currently marked by the user. A posterior probability  $P(u|d,q,con)$  that the document is of interest to the user, given a search query submitted during a current interaction session, can also be calculated.

A variety of personalized information services are provided using the estimated probabilities. In one application, network documents are crawled and parsed for links, and probable interest of the user in the links is calculated using the learning machine. Links likely to be of interest to the user are followed. In another application, the user identifies a document, and a score derived from the estimated probability is provided to the user. In an additional application, the user is provided with a three-dimensional map indicating user interest in each document of a hyperlinked document collection. In a further application, an expert user is selected from a group of users. The expert user has an expert User Model that indicates a strong interest in a document associated with a

particular area of expertise. Another application includes parsing a viewed document for hyperlinks and separately estimating for each hyperlink a probability that the linked document is of interest to the user. In a further application, user interest information derived from the User Model is sent to a third party web server that then customizes its interaction with the user. Finally, a set of users interested in a document is identified, and a range of interests for the identified users is calculated.

## BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 is a schematic diagram of a computer system in which the present invention is implemented.

Fig. 2 is a block diagram of a method of the present invention for providing personalized product and information services to a user.

Fig. 3 is a schematic diagram of knowledge sources used as inputs to the User Model and resulting outputs.

Figs. 4A-4E illustrate tables that store different components and parameters of the User Model.

Fig. 5A illustrates a cluster tree containing clusters of users similar to a particular user.

Fig. 5B is a table that stores parameters of a user cluster tree.

Fig. 6A illustrates a preferred cluster tree for implementing fuzzy or probabilistic clustering.

Fig. 6B is a table that stores parameters of a user fuzzy cluster tree.

Fig. 7 illustrates a portion of a topic tree.

Fig. 8 is a table that stores nodes of the topic tree of Fig. 7.

Fig. 9 is a table that stores the names of clusters having the most interest in nodes of the topic tree of Fig. 7, used to implement the topic experts model.

Fig. 10 illustrates a portion of a product tree.

Fig. 11 is a table that stores nodes of the product tree of Fig. 10.

Fig. 12A is a table that stores feature values of products of the product tree of Fig. 10.

Fig. 12B is a table that stores potential values of product features associated with intermediate nodes of the product tree of Fig. 10.

Fig. 13 is a schematic diagram of the method of initializing the User Model.

Fig. 14 illustrates the user recently accessed buffer, which records all user interactions with documents.

Fig. 15A is a table for storing sites that are candidates to include in the user site distribution.

Fig. **15B** is a table for storing words that are candidates to include in the user word distribution.

Fig. **16** is a table that records all products the user has purchased.

Fig. **17** is a schematic diagram of the method of applying the User Model to new documents to estimate the probability of user interest in the document.

Fig. **18** is a block diagram of the personal crawler application of the present invention.

Fig. **19** is a block diagram of the personal search application of the present invention.

Fig. **20** is a block diagram of the personal navigation application of the present invention.

Fig. **21** is a block diagram of the document barometer application of the present invention.

Fig. **22** is a schematic diagram of the three-dimensional map application of the present invention.

### DETAILED DESCRIPTION

Although the following detailed description contains many specifics for the purposes of illustration, anyone of ordinary skill in the art will appreciate that many variations and alterations to the following details are within the scope of the invention. Accordingly, the following preferred embodiment of the invention is set forth without any loss of generality to, and without imposing limitations upon, the claimed invention.

The present invention, referred to as Personal Web, provides automatic, personalized information and product services to a computer network user. In particular, Personal Web is a user-controlled, web-centric service that creates for each user a personalized perspective and the ability to find and connect with information on the Internet, in computer networks, and from human experts that best matches his or her interests and needs. A computer system **10** implementing Personal Web **12** is illustrated schematically in **Fig. 1**. Personal Web **12** is stored on a central computer or server **14** on a computer network, in this case the Internet **16**, and interacts with client machines **18, 20, 22, 24, 26** via client-side software. Personal Web **12** may also be stored on more than one central computers or servers that interact over the network. The client-side software may be part of a web browser, such as Netscape Navigator or Microsoft Internet Explorer, configured to interact with Personal Web **12**, or it may be distinct from but interacting with a client browser. Five client machines are illustrated for simplicity, but Personal Web **12** is intended to provide personalized web services for a large number of clients simultaneously.

For all of the typical interactions that a user has with a computer network, such as the world wide web, Personal Web 12 provides a personalized version. Personal Web 12 stores for each user a User Model 13 that is continuously and transparently updated based on the user's interaction with the network, and which allows for personalization of all interaction modes. The User Model represents the user's information and product interests; all information that is presented to the user has been evaluated by the User Model to be of interest to the user. The User Model allows for cross fertilization; that is, information that is learned in one mode of interaction is used to improve performance in all modes of interaction. The User Model is described in detail below.

Five examples of personalized interaction modes provided by the present invention are illustrated in Fig. 1. However, it is to be understood that the present invention provides for personalization of all modes, and that the following examples in no way limit the scope of the present invention. Personal Web is active during all stages of information processing, including collection, retrieval, filtering, routing, and query answering.

Client 18 performs a search using Personal Web 12 by submitting a query and receiving personalized search results. The personal search feature collects, indexes, and filters documents, and responds to the user query, all based on the user profile stored in the User Model 13. For example, the same query (e.g., "football game this weekend" or "opera") submitted by a teenager in London and an adult venture capitalist in Menlo Park returns different results based on the personality, interests, and demographics of each user. By personalizing the collection phase, the present invention does not require that all network documents be filtered for a particular user, as does the prior art.

Client 20 browses the web aided by Personal Web 12. In browsing mode, the contents of a web site are customized according to the User Model 13. Personal Web interacts with a cooperating web site by supplying User Model information, and a web page authored in a dynamic language (e.g., DHTML) is personalized to the user's profile. In navigation mode, a personal navigation aid suggests to the user relevant links within the visited site or outside it given the context, for example, the current web page and previously visited pages, and knowledge of the user profile.

Client 22 illustrates the find-an-expert mode of Personal Web 12. The user supplies an expert information or product need in the form of a sample web page or text string, and Personal Web 12 locates an expert in the user's company, circle of friends, or outside



groups that has the relevant information and expertise, based on the expert's User Model **13**. The located expert not only has the correct information, but presents it in a manner of most interest to the user, for example, focussing on technical rather than business details of a product.

Client **24** uses the personal pushed information mode of Personal Web **12**. Personal Web **12** collects and presents personal information to a user based on the User Model **13**. The pushed information is not limited to a fixed or category or topic, but includes any information of interest to the user. In communities, organizations, or group of users, the pushed information can include automatic routing and delivery of newly created documents that are relevant to the users.

Finally, client **26** illustrates the product recommendation mode of Personal Web **12**. The user submits a query for information about a product type, and Personal Web **12** locates the products and related information that are most relevant to the user, based on the User Model **13**. As described further below, product information is gathered from all available knowledge sources, such as product reviews and press releases, and Personal Web **12** can recommend a product that has never been purchased or rated by any users.

All of the above features of Personal Web **12** are based on a User Model **13** that represents user interests in a document or product independently of any specific user information need, i.e., not related to a specific query. The User Model **13** is a function that is developed and updated using a variety of knowledge sources and that is independent of a specific representation or data structure. The underlying mathematical framework of the modeling and training algorithms discussed below is based on Bayesian statistics, and in particular on the optimization criterion of maximizing posterior probabilities. In this approach, the User Model is updated based on both positive and negative training examples. For example, a search result at the top of the list that is not visited by the user is a negative training example.

The User Model **13**, with its associated representations, is an implementation of a learning machine. As defined in the art, a learning machine contains tunable parameters that are altered based on past experience. Personal Web **12** stores parameters that define a User Model **13** for each user, and the parameters are continually updated based on monitored user interactions while the user is engaged in normal use of a computer. While

a specific embodiment of the learning machine is discussed below, it is to be understood that any model that is a learning machine is within the scope of the present invention.

The present invention can be considered to operate in three different modes: initialization, updating or dynamic learning, and application. In the initialization mode, a User Model 13 is developed or trained based in part on a set of user-specific documents. The remaining two modes are illustrated in the block diagram of Fig. 2. While the user is engaged in normal use of a computer, Personal Web 12 operates in the dynamic learning mode to transparently monitor user interactions with data (step 30) and update the User Model 13 to reflect the user's current interests and needs. This updating is performed by updating a set of user-specific data files in step 32, and then using the data files to update the parameters of the User Model 13 in step 34. The user-specific data files include a set of documents and products associated with the user, and monitored user interactions with data. Finally, Personal Web 12 applies the User Model 13 to unseen documents, which are first analyzed in step 36, to determine the user's interest in the document (step 38), and performs a variety of services based on the predicted user interest (step 40). In response to the services provided, the user performs a series of actions, and these actions are in turn monitored to further update the User Model 13.

The following notation is used in describing the present invention. The user and his or her associated representation are denoted with  $u$ , a user query with  $q$ , a document with  $d$ , a product or service with  $p$ , a web site with  $s$ , topic with  $t$ , and a term, meaning a word or phrase, with  $w$ . The term "document" includes not just text, but any type of media, including, but not limited to, hypertext, database, spreadsheet, image, sound, and video. A single document may have one or multiple distinct media types. Accordingly, the set of all possible documents is  $D$ , the set of all users and groups is  $U$ , the set of all products and services is  $P$ , etc. The user information or product need is a subset of  $D$  or  $P$ . Probability is denoted with  $P$ , and a cluster of users or of clusters with  $c$ , with which function semantics are used. For example,  $c(c(u))$  is the cluster of clusters in which the user  $u$  is a member ("the grandfather cluster"). Note that an explicit notation of world knowledge, such as dictionaries, atlases, and other general knowledge sources, which can be used to estimate the various posterior probabilities, is omitted.

A document classifier is a function whose domain is any document, as defined above, and whose range is the continuous interval  $[0,1]$ . For example, a document classifier may be a probability that a document  $d$  is of interest to a particular user or a group of users.

Specific document classifiers of the present invention are obtained using the User Model 13 and Group Model. The User Model 13 represents the user interest in a document independent of any specific user information need. This estimation is unique to each user. In strict mathematical terms, given a user  $u$  and a document  $d$ , the User Model 13 estimates the probability  $P(u|d)$ .  $P(u|d)$  is the probability of the event that the user  $u$  is interested in the document  $d$ , given everything that is known about the document  $d$ . This classifier is extended to include  $P(u|d,con)$ , the probability that a user is interested in a given document based on a user's current context, for example, the web pages visited during a current interaction session.

The Group or Cluster Model is a function that represents the interest level of a group of users in a document independently of any specific information need. For example, for the group of users  $c(u)$ , the mathematical notation of this probability, which is determined by applying the Group Model to a document  $d$ , is  $P(c(u)|d)$ .

A schematic diagram of the User Model is shown in Fig. 3, which illustrates the various knowledge sources (in circles) used as input to the User Model. The knowledge sources are used to initialize and update the User Model, so that it can accurately take documents and generate values of user interest in the documents, given the context of the user interaction. Note that some of the knowledge sources are at the individual user level, while others refer to aggregated data from a group of users, while still others are independent of all users. Also illustrated in Fig. 3 is the ability of the User Model to estimate a user interest in a given product, represented mathematically as the interest of a user in a particular document, given that the document describes the product:  $P(user|document, product\ described = p)$ . As explained further below, the long-term user interest in a product is one of many probabilities incorporated into the computation of user interest in all documents, but it can also be incorporated into estimation of a current user interest in a product.

Beginning at the bottom left of Fig. 3, User Data and Actions include all user-dependent inputs to the User Model, including user browser documents, user-supplied documents, other user-supplied data, and user actions, such as browsing, searching, shopping, finding experts, and reading news. Data and actions of similar users are also incorporated into the User Model by clustering all users into a tree of clusters. Clustering users allows estimation of user interests based on the interests of users similar to the user. For example, if the user suddenly searches for information in an area that is new to him or

her, the User Model borrows characteristics of User Models of users with similar interests. Topic classifiers are used to classify documents automatically into topics according to a predefined topic tree. Similarly, product models determine the product or product categories, if any, referred to by a document. Product models also extract relevant feature of products from product-related documents. The topic experts input provides input of users with a high interest in a particular topic, as measured by their individual User Models. Finally, the User Model incorporates world knowledge sources that are independent of all users, such as databases of company names, yellow pages, thesauri, dictionaries, and atlases.

### **User Model Representations**

Given the inputs shown in Fig. 3, the User Model is a function that may be implemented with any desired data structure and that is not tied to any specific data structure or representation. The following currently preferred embodiment of abstract data structures that represent the User Model **13** is intended to illustrate, but not limit, the User Model of the present invention. Some of the structures hold data and knowledge at the level of individual users, while others store aggregated data for a group or cluster of users. Initialization of the various data structures of the User Model is described in the following section; the description below is of the structures themselves.

User-dependent inputs are represented by components of the User Model shown in **Figs. 4A-4E**. These inputs are shown as tables for illustration purposes, but may be any suitable data structure. The user-dependent components include an informative word or phrase list, a web site distribution, a user topic distribution, a user product distribution, and a user product feature distribution. Each of these user-dependent data structures can be thought of as a vector of most informative or most frequent instances, along with a measure representing its importance to the user.

The informative word and phrase list of Fig. 4A contains the most informative words and phrases found in user documents, along with a measure of each informative phrase or word's importance to the user. As used herein, an "informative phrase" includes groups of words that are not contiguous, but that appear together within a window of a predefined number of words. For example, if a user is interested in the 1999 Melissa computer virus, then the informative phrase might include the words "virus," "Melissa," "security," and "IT," all appearing within a window of 50 words. The sentence "The

computer virus Melissa changed the security policy of many IT departments” corresponds to this phrase.

In addition to the words and phrases, the list contains the last access time of a document containing each word or phrase and the total number of accessed documents containing the words. One embodiment of the informative measure is a word probability distribution  $P(w|u)$  representing the interest of a user  $u$  in a word or phrase  $w$ , as measured by the word’s frequency in user documents. Preferably, however, the informative measure is not simply a measure of the word frequency in user documents; common words found in many documents, such as “Internet,” provide little information about the particular user’s interest. Rather, the informative measure should be high for words that do not appear frequently across the entire set of documents, but whose appearance indicates a strong likelihood of the user’s interest in a document. A preferred embodiment uses the TFIDF measure, described in Ricardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999, in which TF stands for term frequency, and IDF stands for inverse document frequency. Mathematically, if  $f_{u,w}$  denotes the frequency of the word  $w$  in user  $u$  documents, and  $D_w$  denotes the number of documents containing the word  $w$ , then the importance of a word  $w$  to a user  $u$  is proportional to the product  $f_{u,w} \cdot D/D_w$ .

A more preferred embodiment of the measure of each word’s importance uses a mathematically sound and novel implementation based on information theory principles. In particular, the measure used is the mutual information between two random variables representing the user and the word or phrase. Mutual information is a measure of the amount of information one random variable contains about another; a high degree of mutual information between two random variables implies that knowledge of one random variable reduces the uncertainty in the other random variable.

For the present invention, the concept of mutual information is adapted to apply to probability distributions on words and documents. Assume that there is a document in which the user’s interest must be ascertained. The following two questions can be asked: Does the phrase  $p$  appear in the document?; and Is the document of interest to the user  $u$ ? Intuitively, knowing the answer to one of the questions reduces the uncertainty in answering the other question. That is, if the word  $w$  appears in a different frequency in the documents associated with the user  $u$  from its frequency in other documents, it helps reduce the uncertainty in determining the interest of user  $u$  in the document.

Through the concept of mutual information, information theory provides the mathematical tools to quantify this intuition in a sound way. For a detailed explanation, see T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991. In this embodiment of the informative measure, two indicator variables are defined.  $I_w$  has a value of 1 when the word  $w$  appears in a web document and 0 when it does not, and  $I_u$  has a value of 1 when a web document is of interest to the user  $u$  and 0 when it does not. The mutual information between the two random variables  $I_w$  and  $I_u$  is defined as:

$$I(I_w; I_u) = \sum_{i_w \in I_w} \sum_{i_u \in I_u} P(i_w, i_u) \log_2 \frac{P(i_w, i_u)}{P(i_w)P(i_u)}$$

The probabilities in this formula are computed over a set of documents of interest to the user and a set of documents not of interest to the user. For example, consider a set of 100 documents of interest to the user, and a set of 900 documents not of interest to the user. Then  $P(i_u=1) = 0.1$ , and  $P(i_u=0) = 0.9$ . Assume that in the combined set of 1000 documents, 150 contain the word “Bob.” Then  $P(i_w=1) = 0.15$ , and  $P(i_w=0) = 0.85$ . In addition, assume that “Bob” appears in all 100 of the documents of interest to the user.  $P(i_w, i_u)$  has the following four values:

$i_u$	$i_w$	$P(i_w, i_u)$
0	0	850/1000
0	1	50/1000
1	0	0/1000
1	1	100/1000

Using the above formula, the mutual information between the user and word Bob is:

$$\begin{aligned} I(I_{Bob}; I_{user}) &= 850/1000 \log [850/1000 / (0.85 * 0.9)] + 50/1000 \log [50/1000 / (0.15 * 0.9)] \\ &\quad + 0/1000 \log [0/1000 / (0.1 * 0.85)] + 100/1000 \log [100/100 / (0.15 * 0.1)] \\ &= 0.16. \end{aligned}$$

Mutual information is a preferred measure for selecting the word and phrase list for each user. The chosen words and phrases have the highest mutual information.

The remaining User Model representations are analogously defined using probability distributions or mutual information. The web site distribution of Fig. 4B contains a list of

web sites favored by the user along with a measure of the importance of each site. Given the dynamic nature of the Internet, in which individual documents are constantly being added and deleted, a site is defined through the first backslash (after the www). For example, the uniform resource locator (URL) <http://www.herring.com/companies/2000...> is considered as [www.herring.com](http://www.herring.com). Sites are truncated unless a specific area within a site is considered a separate site; for example, [www.cnn.com/health](http://www.cnn.com/health) is considered to be a different site than [www.cnn.com/us](http://www.cnn.com/us). Such special cases are decided experimentally based on the amount of data available on each site and the principles of data-driven approaches, described in Vladimir S. Cherkassky and Filip M. Mulier, *Learning from Data: Concepts, Theory, and Methods*, in *Adaptive and Learning Systems for Signal Processing, Communications and Control*, Simon Haykin, series editor, Wiley & Sons, March, 1998. Each site has an importance measure, either a discrete probability distribution,  $P(s|u)$ , representing the interest of user  $u$  in a web site  $s$ , or the mutual information metric defined above,  $I(I_s; I_u)$ , representing the mutual information between the user  $u$  and a site  $s$ . The web site distribution also contains the last access time and number of accesses for each site.

Fig. 4C illustrates the user topic distribution, which represents the interests of the user in various topics. The user topic distribution is determined from a hierarchical, user-independent topic model, for example a topic tree such as the Yahoo directory or the Open Directory Project, available at <http://dmoz.org/>. Each entry in the tree has the following form:

Computers\Internet\WWW\Searching the Web\Directories\Open Directory Project\

where the topic following a backslash is a child node of the topic preceding the backslash. The topic model is discussed in more detail below.

For each node of the topic tree, a probability is defined that specifies the user interest in the topic. Each level of the topic model is treated distinctly. For example, for the top level of the topic model, there is a distribution in which

$$P(t_u | u) + P(t_l | u) = 1,$$

where  $t_l$  represents the top level of topics and is the same set of topics for each user, e.g., technology, business, health, etc.  $P(t_l | u)$  is the sum of the user probabilities on all top level topics. For each topic level,  $t_u$  represents specific interests of each user that are not part of any common interest topics, for instance family and friends' home pages. For

lower topic levels, every node in the tree is represented in the user topic distribution by a conditional probability distribution. For example, if the Technology node splits into Internet, Communication, and Semiconductors, then the probability distribution is of the form:

$$P(\text{Internet} | u, \text{Technology}) + P(\text{Communication} | u, \text{Technology}) + \\ P(\text{Semiconductors} | u, \text{Technology}) + P(t_u | u, \text{Technology}) = 1$$

Rather than probabilities, the mutual information metric defined above may be used;  $I(I_i; I_u)$  represents the mutual information between the user  $u$  and the topic  $t$ . An exemplary data structure shown in Fig. 4C for storing the user topic distribution contains, for each topic, the topic parent node, informative measure, last access time of documents classified into the topic, and number of accesses of documents classified into the topic. Note that the User Model contains an entry for every topic in the tree, some of which have a user probability or mutual information of zero.

The user product distribution of Fig. 4D represents the interests of the user in various products, organized in a hierarchical, user-independent structure such as a tree, in which individual products are located at the leaf nodes of the tree. The product taxonomy is described in further detail below. The product taxonomy is similar to the topic tree. Each entry in the tree has the following form:

Consumer Electronics\Cameras\Webcams\3Com HomeConnect\

where a product or product category following a backslash is a child node of a product category preceding the backslash.

For each node of the product model, a probability is defined that specifies the user interest in that particular product or product category. Each level of the product model is treated distinctly. For example, for the top level of the product hierarchy, there is a distribution in which

$$P(p_i | u) = 1,$$

where  $p_i$  represents the top level of product categories and is the same for each user, e.g., consumer electronics, computers, software, etc. For lower product category levels, every node in the tree is represented in the user product distribution by a conditional probability distribution. For example, if the Cameras node splits into Webcams and Digital Cameras, then the probability distribution is of the form:

$$P(\text{Webcams} | u, \text{Cameras}) + P(\text{Digital Cameras} | u, \text{Cameras}) = 1$$

Rather than probabilities, the mutual information metric defined above may be used. Then  $I(I_p; I_u)$  represents the mutual information between the user  $u$  and the product or



product category  $p$ . An exemplary data structure for storing the user product distribution contains, for each product, the product ID, product parent node, user probability, last purchase time of the product, number of product purchases, last access time of documents related to the product, and number of related documents accessed.

For each product or category on which the user has a nonzero probability, the User Model contains a user product feature distribution on the relevant features, as shown in Fig. 4E. Each product category has associated with it a list of features, and the particular values relevant to the user are stored along with a measure of the value's importance, such as a probability  $P(f|u,p)$  or mutual information measure  $I(I_f; I_w)$ . For example, Webcams have a feature Interface with possible values Ethernet (10BaseT), Parallel, PC Card, serial, USB, and TV. Probability values of each feature sum to one; that is,

$$P(\text{Ethernet} | u, \text{Interface}, \text{Webcam}) + P(\text{Parallel} | u, \text{Interface}, \text{Webcam}) + P(\text{PC Card} | u, \text{Interface}, \text{Webcam}) + P(\text{serial} | u, \text{Interface}, \text{Webcam}) + P(\text{USB} | u, \text{Interface}, \text{Webcam}) + P(\text{TV} | u, \text{Interface}, \text{Webcam}) = 1.$$

User probability distributions or mutual information measures are stored for each feature value of each node. Note that there is no user feature value distribution at the leaf nodes, since specific products have particular values of each feature.

Finally, user-dependent components of the User Model include clusters of users similar to the user. Users are clustered into groups, forming a cluster tree. One embodiment of a user cluster tree, shown in Fig. 5A, hard classifies users into clusters that are further clustered. Each user is a member of one and only one cluster. For example, Bob is clustered into a cluster  $c(u)$ , which is further clustered into clusters of clusters, until the top level cluster is reached  $c(U)$ . The identity of the user's parent cluster and grandfather cluster is stored as shown in Fig. 5B, and information about the parent cluster is used as input into the User Model. As described below, clusters are computed directly from User Models, and thus need not have a predefined semantic underpinning.

Preferably, the User Model does not use hard clustering, but rather uses soft or fuzzy clustering, also known as probabilistic clustering, in which the user belongs to more than one cluster according to a user cluster distribution  $P(c(u))$ . Fig. 6A illustrates fuzzy clusters in a cluster hierarchy. In this case, Bob belongs to four different clusters according to the probability distribution shown. Thus Bob is most like the members of

cluster C4, but still quite similar to members of clusters C1, C2, C3, and C4. Fuzzy clustering is useful for capturing different interests of a user. For example, a user may be a small business owner, a parent of a small child, and also an avid mountain biker, and therefore need information for all three roles. Probabilistic clustering is described in detail in the Ph.D. thesis of Steven J. Nowlan, "Soft Competitive Adaptation: Neural Network Learning Algorithms Based on Fitting Statistical Mixtures," School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1991. A suitable data structure for representing fuzzy clusters is shown in Fig. 6B. Each row stores the cluster or user ID, one parent ID, and the cluster probability, a measure of similarity between the cluster or user and the parent cluster.

Note that all elements of an individual User Model for a user  $u$  also apply to a cluster of users  $c(u)$ . Thus for each cluster, a Group Model is stored containing an informative word list, a site distribution, a topic distribution, a group product distribution, and a group product feature distribution, each with appropriate measures. For example,  $P(p|c(u))$  represents the interest of a cluster  $c(u)$  in various products  $p$ .

The user-dependent User Model representations also include a user general information table, which records global information describing the user, such as the User ID, the number of global accesses, the number of accesses within a recent time period, and pointers to all user data structures.

Other knowledge sources of the User Model are independent of the user and all other users. Topic classifiers are used to classify documents into topics according to a predefined topic tree, an example of which is illustrated in Fig. 7. A variety of topic trees are available on the web, such as the Yahoo directory or Open Directory Project ([www.dmoz.org](http://www.dmoz.org)). A topic classifier is a model similar to the user model that estimates the probability that a document belongs to a topic. Every node on the topic tree has a stored topic classifier. Thus the set of all topic classifiers computes a probability distribution of all of the documents in the set of documents  $D$  among the topic nodes. For example, the topic classifier in the root node in Fig. 7 estimates the posterior probabilities  $P(t|d)$ , where  $t$  represents the topic of document  $d$  and is assigned values from the set {Arts, Business, Health, News, Science, Society}. Similarly, the topic classifier for the Business node estimates the posterior probability  $P(t|d, \text{Business})$ , where  $t$  represents the specific topic of the document  $d$  within the Business category. Mathematically, this posterior probability is denoted  $P(t(d) = \text{Business} \setminus \text{Investing} \mid t(d) = \text{Business}, d)$ , which

represents the probability that the subtopic of the document  $d$  within Business is Investing, given that the topic is Business. The topic tree is stored as shown in Fig. 8, a table containing, for each node, the topic ID, depth level, topic parent ID, number of child nodes, and topic ID of the child nodes.

The topic experts model estimates the probability that a document is of interest to users who are interested in a particular topic, independent of any specific user information need. Each node of the topic tree has, in addition to a topic classifier, a corresponding topic expert function. Note that the topic classifier and topic expert function are independent; two documents can be about investing, but one of high interest to expert users and the other of no interest to expert users. The topic expert model can be considered an evaluation of the quality of information in a given document. The assumption behind the topic experts model is that the degree of interest of a user in a given topic is his or her weight for predicting the quality or general interest level in a document classified within the particular topic. Obviously there are outliers to this assumption, for example, novice users. However, in general and averaged across many users, this measure is a good indicator of a general interest level in a document. For every topic in the tree, a list of the  $N$  clusters with the most interest in the topic based on the cluster topic distribution is stored. The cluster topic distribution is similar to the user topic distribution described above, but is averaged over all users in the cluster. An exemplary data structure for storing the topic experts model is shown in Fig. 9.

Finally, a product model is stored for every node of a product taxonomy tree, illustrated in Fig. 10. Examples of product taxonomy trees can be found at [www.cnet.com](http://www.cnet.com) and [www.productopia.com](http://www.productopia.com), among other locations. In any product taxonomy tree, the leaf nodes, i.e., the bottom nodes of the tree, correspond to particular products, while higher nodes represent product categories. Product models are similar to topic classifiers and User Models, and are used to determine whether a document is relevant to a particular product or product category. Thus a product model contains a list of informative words, topics, and sites. The set of all product models computes a probability distribution of all of the documents in the set of documents  $D$  among the product nodes. For example, the product model in the root node in Fig. 10 estimates the posterior probabilities  $P(p | d)$ , where  $p$  represents the product referred to in document  $d$  and is assigned values from the set {Consumer Electronics, Computers, Software}. Similarly, the product model for the Consumer Electronics node estimates the posterior probability  $P(p | d, \text{Consumer Electronics})$ , where  $p$  represents the product category of the document  $d$  within the

Consumer Electronics category. Mathematically, this posterior probability is denoted  $P(p(d) = \text{Consumer Electronics} \setminus \text{CD Players} \mid p(d) = \text{Consumer Electronics}, d)$ , which represents the probability that the subproduct category of the document  $d$  within Consumer Electronics is CD Players, given that the product category is Consumer Electronics. The product tree is stored as shown in Fig. 11, a table containing, for each node, the topic ID, depth level, topic parent ID, number of child nodes, and topic ID of the child nodes.

Each node of the product tree has an associated product feature list, which contains particular descriptive features relevant to the product or category. Nodes may have associated feature values; leaf nodes, which represent specific products, have values of all relevant product features. Product feature lists are determined by a human with knowledge of the domain. However, feature values may be determined automatically from relevant knowledge sources as explained below.

For example, in the product tree of Fig. 10, CD Players is the parent node of the particular CD players Sony CDP-CX350 and Harman Kardon CDR2. The product category CD Players has the following features: Brand, CD Capacity, Digital Output, Plays Minidisc, and Price Range. Each feature has a finite number of potential feature values; for example, CD Capacity has potential feature values 1 Disc, 1-10 Discs, 10-50 Discs, or 50 Discs or Greater. Individual products, the child nodes of CD Players, have one value of each feature. For example, the Sony CDP-CX350 has a 300 disc capacity, and thus a feature value of 50 Discs or Greater.

Some product features are relevant to multiple product categories. In this case, product features propagate as high up the product tree as possible. For example, digital cameras have the following product features: PC Compatibility, Macintosh Compatibility, Interfaces, Viewfinder Type, and Price Range. Webcams have the following product features: PC Compatibility, Macintosh Compatibility, Interfaces, Maximum Frames per Second, and Price Range. Common features are stored at the highest possible node of the tree; thus features PC Compatibility, Macintosh Compatibility, and Interfaces are stored at the Cameras node. The Digital Cameras node stores only product feature Viewfinder Type, and the Webcams node stores only product feature Maximum Frames per Second. Note that product feature Price Range is common to CD Players and Cameras, and also Personal Minidiscs, and thus is propagated up the tree and stored at node Consumer Electronics.

Individual products at leaf nodes inherit relevant features from all of their ancestor nodes. For example, Kodak CD280 inherits the feature Viewfinder Type from its parent; PC Compatibility, Macintosh Compatibility, and Interfaces from its grandparent; and Price Range from its great-grandparent. A product feature list is stored as shown in **Fig. 12A**, and contains, for each product ID, the associated feature and its value. All potential feature values are stored in a product feature value list, as shown in **Fig. 12B**.

The system also includes a document database that indexes all documents  $D$ . The document database records, for each document, a document ID, the full location (the URL of the document), a pointer to data extracted from the document, and the last access time of the document by any user. A word database contains statistics of each word or phrase from all user documents. The word database contains the word ID, full word, and word frequency in all documents  $D$ , used in calculating informative measures for individual users and clusters.

### **Initialization of User Model**

The User Model is initialized offline using characterizations of user behavior and/or a set of documents associated with the user. Each data structure described above is created during initialization. In other words, the relevant parameters of the learning machine are determined during initialization, and then continually updated online during the update mode.

In one embodiment, the user documents for initializing the User Model are identified by the user's web browser. Most browsers contain files that store user information and are used to minimize network access. In Internet Explorer, these files are known as favorites, cache, and history files. Most commercial browsers, such as Netscape Navigator, have equivalent functionality; for example, bookmarks are equivalent to favorites. Users denote frequently-accessed documents as bookmarks, allowing them to be retrieved simply by selection from the list of bookmarks. The bookmarks file includes for each listing its creation time, last modification time, last visit time, and other information. Bookmarks of documents that have changed since the last user access are preferably deleted from the set of user documents. The Internet Temporary folder contains all of the web pages that the user has opened recently (e.g., within the last 30 days). When a user views a web page, it is copied to this folder and recorded in the cache file, which contains the following fields: location (URL), first access time, and last access time (most recent

retrieval from cache). Finally, the history file contains links to all pages that the user has opened within a set time period.

Alternatively, the user supplies a set of documents, not included in any browser files, that represent his or her interests. The User Model can also be initialized from information provided directly by the user. Users may fill out forms, answer questions, or play games that ascertain user interests and preferences. The user may also rate his or her interest in a set of documents provided.

User documents are analyzed as shown in **Fig. 13** to determine initial parameters for the various functions of the User Model. A similar analysis is used during updating of the User Model. Note that during updating, both documents that are of interest to the user and documents that are not of interest to the user are analyzed and incorporated into the User Model. The process is as follows. In a first step **82**, the format of documents **80** is identified. In step **84**, documents **80** are parsed and separated into text, images and other non-text media **88**, and formatting. Further processing is applied to the text, such as stemming and tokenization to obtain a set of words and phrases **86**, and information extraction. Through information extraction, links **90** to other documents, email addresses, monetary sums, people's names, and company names are obtained. Processing is performed using natural language processing tools such as LinguistX<sup>®</sup> and keyword extraction tools such as Thing Finder<sup>™</sup>, both produced by Inxight ([www.inxight.com](http://www.inxight.com)). Further information on processing techniques can be found in Christopher D. Manning and Hinrich Schutze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999. Additional processing is applied to images and other non-text media **88**. For example, pattern recognition software determines the content of images, and audio or speech recognition software determines the content of audio. Finally, document locations **94** are obtained.

Parsed portions of the documents and extracted information are processed to initialize or update the user representations in the User Model. In step **96**, user informative words or phrases **98** are obtained from document words and phrases **86**. In one embodiment, a frequency distribution is obtained to calculate a TFIDF measure quantifying user interest in words **98**. Alternatively, mutual information is calculated between the two indicator variables  $I_w$  and  $I_d$  as explained above. The set of informative words **98** contains words with the highest probabilities or mutual information.

In step **100**, the topic classifiers are applied to all extracted information and portions of documents **80** to obtain a probability distribution  $P(t|d)$  for each document on each node of the topic tree. As a result, each node has a set of probabilities, one for each document, which is averaged to obtain an overall topic node probability. The average probabilities become the initial user topic distribution **102**. If desired, mutual information between the two indicator variables  $I_s$  and  $I_u$  can be determined as explained above.

Similarly, in step **104**, product models are applied to all extracted information from documents **80** to classify documents according to the product taxonomy tree. From user purchase history **105**, additional product probabilities are obtained. Probabilities for each node are combined, weighting purchases and product-related documents appropriately, to obtain a user product distribution **106**. Note that only some of documents **80** contain product-relevant information and are used to determine the user product distribution **106**. Product models return probabilities of zero for documents that are not product related.

The user product feature distribution **108** can be obtained from different sources. If a user has a nonzero probability for a particular product node, then the feature distribution on that node is obtained from its leaf nodes. For example, if one of the user documents was classified into Kodak DC280 and another into Nikon Coolpix 950, then the user product feature distribution for the Digital Cameras node has a probability of 0.5 for the feature values corresponding to each camera. Feature value distributions propagate throughout the user product feature distributions. For example, if the two cameras are in the same price range, \$300-\$400, then the probability of the value \$300-\$400 of the feature Price Range is 1.0, which propagates up to the Consumer Electronics node (assuming that the user has no other product-related documents falling within Consumer Electronics).

Alternatively, product feature value distributions are obtained only from products that the user has purchased, and not from product-related documents in the set of user documents. Relevant feature values are distributed as high up the tree as appropriate. If the user has not purchased a product characterized by a particular feature, then that feature has a zero probability. Alternatively, the user may explicitly specify his or her preferred feature values for each product category in the user product distribution. User-supplied information may also be combined with feature value distributions obtained from documents or purchases.

Document locations **94** are analyzed (step **110**) to obtain the user site distribution **112**. Analysis takes into account the relative frequency of access of the sites within a recent time period, weighted by factors including how recently a site was accessed, whether it was kept in the favorites or bookmarks file, and the number of different pages from a single site that were accessed. Values of weighting factors are optimized experimentally using jackknifing and cross-validation techniques described in H. Bourlard and N. Morgan, *Connectionist Speech Recognition: A Hybrid Approach*, Kluwer Academic Publishers, 1994.

Note that there is typically overlap among the different representations of the User Model. For example, a news document announcing the release of a new generation of Microsoft servers has relevant words Microsoft and server. In addition, it is categorized within the product taxonomy under Microsoft servers and the topic taxonomy under computer hardware. This document may affect the user's word list, product distribution, and topic distribution.

After the User Models are initialized for all users, cluster membership can be obtained. Clusters contain users with a high degree of similarity of interests and information needs. A large number of clustering algorithms are available; for examples, see K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, 1990. As discussed above, users are preferably soft clustered into more than one cluster. Preferably, the present invention uses an algorithm based on the relative entropy measure from information theory, a measure of the distance between two probability distributions on the same event space, described in T. Cover and J. Thomas, *Elements of Information Theory*, Chapter 2, Wiley, 1991. Clustering is unsupervised. That is, clusters have no inherent semantic significance; while a cluster might contain users with a high interest in mountain biking, the cluster tree has no knowledge of this fact.

In a preferred embodiment, the relative entropy between two User Model distributions on a fixed set of documents  $D_{sample}$  is calculated.  $D_{sample}$  is chosen as a good representation of the set of all documents  $D$ . Distributions of similar users have low relative entropy, and all pairs of users within a cluster have relative entropy below a threshold value. The User Model of each user is applied to the documents to obtain a probability of interest of each user in each document in the set. The relative entropy between two user distributions for a single document is calculated for each document in the set, and then summed across all documents.



The exact mathematical computation of the relative entropy between two users is as follows. An indicator variable  $I_{u,d}$  is assigned to 1 when a document  $d$  is of interest to a user  $u$  and 0 when it is not. For two users  $u_1$  and  $u_2$  and for any document  $d$ , the relative entropy between the corresponding distributions is:

$$D(I_{u_1,d} \| I_{u_2,d}) = \sum_{i \in I} P(i_{u_1,d}) \log_2 \frac{P(i_{u_1,d})}{P(i_{u_2,d})}$$

For example, if  $P(u_1|d)=0.6$  and  $P(u_2|d)=0.9$ , then

$$D(I_{u_1,d} \| I_{u_2,d}) = 0.4 \log (0.4/0.1) + 0.6 \log (0.6/0.9).$$

The relative entropy can be converted to a metric  $D'$  that obeys the triangle inequality:

$$D'(I_1 \| I_2) = 0.5 * (D(I_1 \| I_2) + D(I_2 \| I_1)).$$

For any two users  $u_1$  and  $u_2$ , and for each document in  $D_{sample}$ , the metric  $D'$  is computed between the corresponding indicator variable distributions on the document. The values for all document are summed, and this sum is the distance metric for clustering users. This distance is defined as:

$$\text{Distance}(u_1, u_2) = \sum_{d_j \in D_{sample}} D'(I_{u_1,d_j} \| I_{u_2,d_j}).$$

An alternative clustering algorithm computes the relative entropy between individual user distributions in the User Model, for example, between all informative word lists, site distributions, etc., of each user. The equations are similar to those above, but compute relative entropy based on indicator variables such as  $I_{u,w}$ , which is assigned a value of 1 when a word  $w$  is of interest to a user  $u$ . The calculated distances between individual user distributions on words, sites, topics, and products are summed to get an overall user distance. This second algorithm is significantly less computationally costly than the preferred algorithm above; selection of an algorithm depends on available computing resources. In either case, relative entropy can also be computed between a user and cluster of users.

Each cluster has a Group or Cluster Model that is analogous to a User Model. Cluster Models are generated by averaging each component of its members' User Models. When fuzzy clusters are used, components are weighted by a user's probability of membership in the cluster.

In some cases, initialization is performed without any user-specific information. A user may not have a large bookmarks file or cache, or may not want to disclose any personal information. For such users, prototype users are supplied. A user can choose one or a combination of several prototype User Models, such as the technologist, the art lover, and the sports fan. Predetermined parameters of the selected prototype user are used to initialize the User Model. Users can also opt to add only some parameters of a prototype user to his or her existing User Model by choosing the prototype user's distribution of topics, words, sites, etc. Note that prototype users, unlike clusters, are semantically meaningful. That is, prototype users are trained on a set of documents selected to represent a particular interest. For this reason, prototype users are known as "hats," as the user is trying on the hat of a prototype user.

Users can also choose profiles on a temporary basis, for a particular session only. For example, in a search for a birthday present for his or her teenage daughter, a venture capitalist from Menlo Park may be interested in information most probably offered to teenagers, and hence may choose a teenage girl profile for the search session.

User-independent components are also initialized. The topic classifiers are trained using the set of all possible documents  $D$ . For example,  $D$  may be the documents classified by the Open Directory Project into its topic tree. Topic classifiers are similar to a User Model, but with a unimodal topic distribution function (i.e., a topic model has a topic distribution value of 1 for itself and 0 for all other topic nodes). The set of documents associated with each leaf node of the topic tree is parsed and analyzed as with the user model to obtain an informative word list and site distribution. When a topic classifier is applied to a new document, the document's words and location are compared with the informative components of the topic classifier to obtain  $P(t|d)$ . This process is further explained below with reference to computation of  $P(u|d)$ . Preferably, intermediate nodes of the tree do not have associated word list and site distributions. Rather, the measures for the word list and site distribution of child nodes are used as input to the topic classifier of their parent nodes. For example, the topic classifier for the Business node of the topic tree of Fig. 7 has as its input the score of the site of the document to be

classified according to the site distributions of the topic models of its child nodes, Employment, Industries, and Investing. The classifier can be any non-linear classifier such as one obtained by training a Multilayer Perceptron (MLP) using jackknifing and cross-validation techniques, as described in H. Bourlard and N. Morgan, *Connectionist Speech Recognition: A Hybrid Approach*, Kluwer Academic Publishers, 1994. It can be shown that a MLP can be trained to estimate posterior probabilities; for details, see J. Hertz, A. Krogh, R. Palmer, *Introduction to The Theory of Neural Computation*, Addison-Wesley, 1991.

The topic experts model is initialized by locating for every node in the topic tree the  $N$  clusters that are of the same depth in the user cluster tree as the user, and that have the highest interest in the topic, based on their cluster topic distribution. The cluster topic distribution  $P(t|c(u))$  is simply an average of the user topic distribution  $P(t|u)$  for each user in the cluster. The topic experts model is used to determine the joint probability that a document and the topic under consideration are of interest to any user,  $P(t, d)$ . Using Bayes' rule, this term can be approximated by considering the users of the  $N$  most relevant clusters.

$$P(t, d) = \sum_{i \in N} P(c_i | t, d) P(t | d) P(d)$$

The topic experts model is, therefore, not a distinct model, but rather an ad hoc combination of user and cluster topic distributions and topic models.

Product models are initialized similarly to User Models and topic classifiers. Each leaf node in the product tree of Fig. 10 has an associated set of documents that have been manually classified according to the product taxonomy. These documents are used to train the product model as shown for the User Model in Fig. 13. As a result, each leaf node of the product tree contains a set of informative words, a topic distribution, and a site distribution. Each node also contains a list of features relevant to that product, which is determined manually. From the documents, values of the relevant features are extracted automatically using information extraction techniques to initialize the feature value list for the product. For example, the value of the CD Capacity is extracted from the document. Information extraction is performed on unstructured text, such as HTML documents, semi-structured text, such as XML documents, and structured text, such as

database tables. As with the topic model, a nonlinear function such as a Multilayer Perceptron is used to train the product model.

Preferably, as for topic classifiers, intermediate nodes of the product tree do not have associated word lists, site distributions, and topic distributions. Rather, the measures for the word list, site distribution, and topic distribution of child nodes are used as input to the product models of their parent nodes. Alternatively, each parent node may be trained using the union of all documents of its child nodes.

### **Updating the User Model**

The User Model is a dynamic entity that is refined and updated based on all user actions. User interactions with network data are transparently monitored while the user is engaged in normal use of his or her computer. Multiple distinct modes of interaction of the user are monitored, including network searching, network navigation, network browsing, email reading, email writing, document writing, viewing pushed information, finding expert advice, product information searching, and product purchasing. As a result of the interactions, the set of user documents and the parameters of each user representation in the User Model are modified.

While any nonlinear function may be used in the User Model (e.g., a Multilayer Perceptron), a key feature of the model is that the parameters are updated based on actual user reactions to documents. The difference between the predicted user interest in a document or product and the actual user interest becomes the optimization criterion for training the model.

Through his or her actions, the user creates positive and negative patterns. Positive examples are documents of interest to a user: search results that are visited following a search query, documents saved in the user favorites or bookmarks file, web sites that the user visits independently of search queries, etc. Negative examples are documents that are not of interest to the user, and include search results that are ignored although appear at the top of the search result, deleted bookmarks, and ignored pushed news or email. Conceptually, positive and negative examples can be viewed as additions to and subtractions from the user data and resources.

Information about each document that the user views is stored in a recently accessed buffer for subsequent analysis. The recently accessed buffer includes information about

the document itself and information about the user's interaction with the document. One possible implementation of a buffer is illustrated in **Fig. 14**; however, any suitable data structure may be used. The recently-accessed buffer contains, for each viewed document, a document identifier (e.g., its URL); the access time of the user interaction with the document; the interaction type, such as search or navigation; the context, such as the search query; and the degree of interest, for example, whether it was positive or negative, saved in the bookmarks file, how long the user spent viewing the document, or whether the user followed any links in the document. Additional information is recorded for different modes of interaction with a document as discussed below.

A metric is determined for each document to indicate whether it is a positive, negative or neutral event; this metric can potentially be any grade between 0 and 1, where 0 is a completely negative event, 1 is a completely positive event, and 0.5 is a neutral event. Previous user interactions may be considered in computing the metric; for example, a web site that the user accesses at a frequency greater than a predetermined threshold frequency is a positive example. After each addition to or subtraction from the set of user documents, the document is parsed and analyzed as for the User Model initialization. Extracted information is incorporated into the User Model.

Because the User Model is constantly and dynamically updated, applying the initialization process for each update is inefficient. Preferably, incremental learning techniques are used to update the User Model. Efficient incremental learning and updating techniques provide for incorporating new items into existing statistics, as long as sufficient statistics are recorded. Details about incremental learning can be found in P. Lee, *Bayesian Statistics*, Oxford University Press, 1989.

After a document stored in the recently accessed buffer is parsed, parsed portions are stored in candidate tables. For example, **Figs. 15A** and **15B** illustrate a user site candidate table and user word candidate table. The user site candidate table holds sites that are candidates to move into the user site distribution of Fig. 4B. The site candidate table stores the site name, i.e., the URL until the first backslash, except for special cases; the number of site accesses; and the time of last access. The user word candidate table holds the words or phrases that are candidates to move into the user informative word list of Fig. 4A. It contains a word or phrase ID, alternate spellings (or misspellings) of the word, an informative grade, and a time of last access.

Negative examples provide words, sites, and topics that can be used in several ways. The measure of any item obtained from the negative example may be reduced in the user distribution. For example, if the negative example is from a particular site that is in the user site distribution, then the probability or mutual information of that site is decreased. Alternatively, a list of informative negative items may be stored. The negative items are obtained from negative examples and are used to reduce the score of a document containing negative items.

Documents are added to the buffer during all user modes of interaction with the computer. Interaction modes include network searching, network navigation, network browsing, email reading, email writing, document writing, viewing "pushed" information, finding expert advice, and product purchasing. Different types of information are stored in the buffer for different modes. In network searching, search queries are recorded and all search results added to the buffer, along with whether or not a link was followed and access time for viewed search results. In network browsing, the user browses among linked documents, and each document is added to the buffer, along with its interaction time. In email reading mode, each piece of email is considered to be a document and is added to the buffer. The type of interaction with the email item, such as deleting, storing, or forwarding, the sender of the email, and the recipient list are recorded. In email writing mode, each piece of written email is considered a document and added to the buffer. The recipient of the email is recorded. Documents written during document writing mode are added to the buffer. The user's access time with each piece of pushed information and type of interaction, such as saving or forwarding, are recorded. In finding expert advice mode, the user's interest in expert advice is recorded; interest may be measured by the interaction time with an email from an expert, a user's direct rating of the quality of information received, or other suitable measure.

During a product purchasing mode, a similar buffer is created for purchased products, as shown in Fig. 16. All purchased products are used to update the User Model. The user recently purchased products buffer records for each purchase the product ID, parent node in the product tree, purchase time, and purchase source. Purchased products are used to update the user product distribution and user product feature distribution.

If the user feels that the User Model is not an adequate representation of him or her, the user may submit user modification requests. For example, the user may request that specific web sites, topics, or phrases be added to or deleted from the User Model.

User Models for prototype users (hats) are also updated based on actions of similar users. Obviously, it is desirable for prototype User Models to reflect the current state of the representative interest. New web sites appear constantly, and even new informative words appear regularly. For example, technology-related words are introduced and widely adopted quite rapidly; the word list of the Technologist hat should be updated to reflect such changes.

Prototype User Models are updated using actions that are related to the prototype. Actions include documents, user reactions to documents, and product purchases. There are many ways to determine whether an action is relevant to the prototype user. A document that is a positive example for many users (i.e., a followed search result or bookmarked page) and also has a high probability of interest to the prototype user is added to the set of prototype user documents. Actions of users or clusters who are similar to the prototype user, as measured by the relative entropy between individual distributions (words, sites, etc.), are incorporated into the prototype User Model. Additions to the prototype User Model may be weighted by the relative entropy between the user performing the action and the prototype user. Actions of expert users who have a high degree of interest in topics also of interest to the prototype user are incorporated into the prototype User Model.

Note that users who are trying on hats are not able to change the prototype User Model. Their actions affect their own User Models, but not the prototype User Model. Updates to the prototype User Model are based only on actions of users who are not currently trying on hats.

Product models are also continually updated using incremental learning techniques. As described below, the present invention includes crawling network documents and evaluating each document against User Models. Crawled documents are also evaluated by product models. Documents that are relevant to a particular product, as determined by the computed probability  $P(p|d)$ , are used to update its product model. If a document is determined to be relevant, then each component of the product model is updated accordingly. In addition to the parsing and analysis performed for user documents, information extraction techniques are employed to derive feature values that are compared against feature values of the product model, and also incorporated into the feature value list as necessary. New products can be added to the product tree at any

time, with characteristic product feature values extracted from all relevant documents. Relevant documents for updating product models include product releases, discussion group entries, product reviews, news articles, or any other type of document.

By employing dynamically updated product models, the present invention, in contrast with prior art systems, provides for deep analysis of all available product information to create a rich representation of products. The interest of a user in a product can therefore be determined even if the product has never been purchased before, or if the product has only been purchased by a very small number of users.

### Applying the User Model to Unseen Documents

The User Model is applied to unseen documents to determine the probability that a document is of interest to the user, or the probability that a document is of interest to a user in a particular context. The basic functionality of this determination is then used in the various applications described in subsequent sections to provide personalized information and product services to the user.

The process of estimating user interest in a particular unseen document **120** is illustrated in **Fig. 17**. This process has the following three steps:

1. Preprocessing the document as for initialization (step **122**).
2. Calculating an individual score for the document for each element of the user representation (e.g., topic distribution, word list).
3. Non-linearly combining (**124**) individual scores into one score **126**, the probability that the user is interested in the unseen document,  $P(u|d)$ .

The second step varies for each individual score. From the parsed text, the words of the document **120** are intersected with the words or phrases in the user informative word list **128**. For every word or phrase in common, the stored mutual information between the two indicator variables  $I_w$  and  $I_u$  is summed to obtain the word score. Alternatively, the TFIDF associated with the word are averaged for every common word or phrase. The location score is given by the probability that the document site is of interest to the user, based on the user site distribution **130**.

The topic classifiers **132** are applied to document **120** to determine the probability that the document relates to a particular topic,  $P(t|d)$ . The user topic score is obtained by computing the relative entropy between the topic distribution  $P(t|d)$  and the user topic



distribution **134**,  $P(t|u)$ . After the document has been classified into topics, the topic expert models **136** are applied as described above to determine a score reflecting the interest of users that are experts in the particular topics of this document.

Similarly, the product models **138** are applied to document **120** to determine which products or product categories it describes,  $P(p|d)$ . From the document product distribution, the product score is obtained by computing the relative entropy between the document product distribution and user product distribution **140**,  $P(p|u)$ . For each product having a nonzero value of  $P(p|d)$ , its feature values are given by the product model. The user's measures on each of these feature values, found in the user product feature distribution **141**, are averaged to obtain a product feature score for each relevant product. Product feature scores are then averaged to obtain an overall product feature score.

The cluster models **142** of clusters to which the user belongs are applied to the document to obtain  $P(c(u)|d)$ . This group model represents the average interests of all users in the cluster. Conceptually, the cluster model is obtained from the union of all the member users' documents and product purchases. Practically, the cluster model is computed from the User Models by averaging the different distributions of the individual User Models, and not from the documents or purchases themselves. Note that in a recursive way, all users have some impact (relative to their similarity to the user under discussion) on the user score, given that  $P(c(u) | d)$  is estimated using  $P(c(c(u))|d)$  as a knowledge source, and so on.

Finally, world knowledge (not shown) is an additional knowledge source that represents the interest of an average user in the document based only on a set of predefined factors. World knowledge factors include facts or knowledge about the document, such as links pointing to and from the document or metadata about the document, for example, its author, publisher, time of publication, age, or language. Also included may be the number of users who have accessed the document, saved it in a favorites list, or been previously interested in the document. World knowledge is represented as a probability between 0 and 1.

In step **124**, all individual scores are combined to obtain a composite user score **126** for document **120**. Step **124** may be performed by training a Multilayer Perceptron using jackknifing and cross-validation techniques, as described in H. Bourlard and N. Morgan,

*Connectionist Speech Recognition: A Hybrid Approach*, Kluwer Academic Publishers, 1994. It has been shown in J. Hertz et al., *Introduction to The Theory of Neural Computation*, Addison-Wesley, 1991, that a Multilayer Perceptron can be trained to estimate posterior probabilities.

The context of a user's interaction can be explicitly represented in calculating the user interest in a document. It is not feasible to update the user model after every newly viewed document or search, but the User Model can be updated effectively instantaneously by incorporating the context of user interactions. Context includes content and location of documents viewed during the current interaction session. For example, if the user visits ten consecutive sites pertaining to computer security, then when the User Model estimates the interest of the user in a document about computer security, it is higher than average. The probability of user interest in a document within the current context  $con$  is given by:

$$P(u|d, con) = \frac{P(u, con|d)}{P(con|d)}$$

In some applications, individual scores that are combined in step 124 are themselves useful. In particular, the probability that a user is interested in a given product can be used to suggest product purchases to a user. If a user has previously purchased a product, then the User Model contains a distribution on the product's features. If these features propagate far up the product tree, then they can be used to estimate the probability that the user is interested in a different type of product characterized by similar features. For example, if the user purchases a digital camera that is Windows compatible, then the high probability of this compatibility feature value propagates up the tree to a higher node. Clearly, all computer-related purchases for this user should be Windows compatible. Every product that is a descendent of the node to which the value propagated can be rated based on its compatibility, and Windows-compatible products have a higher probability of being of interest to the user.

The long-term interest of a user in products, represented by  $P(p|u)$ , is distinct from the user's immediate interest in a product  $p$ , represented as  $P(u|d, product\ described=p)$ . The user's immediate interest is the value used to recommend products to a user. Note that  $P(p|u)$  does not incorporate the user's distribution on feature values. For example, consider the problem of evaluating a user's interest in a particular camera, the Nikon 320.

The user has never read any documents describing the Nikon 320, and so  $P(\text{Nikon } 320|u)=0$ . However, the user's feature distribution for the Cameras node indicates high user interest in all of the feature values characterizing the Nikon 320.

When a given product is evaluated by the User Model, the following measures are combined to obtain  $P(u|d, \text{product described}=p)$ : the probabilities of the product and its ancestor nodes from the user product distribution,  $P(p|u)$ ; an average of probabilities of each feature value from the user product feature distribution,  $P(f|u,p)$ ; a probability from the user's clusters' product distributions,  $P(p|c(u))$ ; and an average of probabilities of feature values from the cluster' product feature distributions,  $P(f|c(u),p)$ . The overall product score is determined by non-linearly combining all measures. The cluster model is particularly useful if the user does not have a feature value distribution on products in which the user's interest is being estimated.

### Applications

The basic function of estimating the probability that a user is interested in a document or product is exploited to provide different types of personalized services to the user. In each type of service, the user's response to the service provided is monitored to obtain positive and negative examples that are used to update the User Model. Example applications are detailed below. However, it is to be understood that all applications employing a trainable User Model as described above are within the scope of the present invention.

### Personal Search

In this application, both the collection and filtering steps of searching are personalized. A set of documents of interest to the user is collected, and then used as part of the domain for subsequent searches. The collected documents may also be used as part of the user documents to update the User Model. The collection step, referred to as Personal Crawler, is illustrated schematically in Fig. 18. A stack 170 is initialized with documents of high interest to the user, such as documents in the bookmarks file or documents specified by the user. If necessary, the stack documents may be selected by rating each document in the general document index according to the User Model. The term "stack" refers to a pushdown stack as described in detail in R. Sedgewick, *Algorithms in C++*, Parts 1-4, Addison-Wesley, 1998.

In step 172, the crawler selects a document from the top of the stack to begin crawling. The document is parsed and analyzed (step 174) to identify any links to other documents. If there are links to other documents, each linked document is scored using the User Model (176). If the linked document is of interest to the user (178), i.e., if  $P(u|d)$  exceeds a threshold level, then it is added to the stack in step 180, and the crawler continues crawling from the linked document (step 172). If the document is not of interest to the user, then the crawler selects the next document on the stack to continue crawling.

The subsequent searching step is illustrated in Fig. 19. In response to a query 190, a set of search results is located from the set containing all documents  $D$  and user documents obtained during personal crawling. The results are evaluated using the User Model (194) and sorted in order of user interest (196), so that the most interesting documents are listed first. The user reaction to each document in the search results is monitored. Monitored reactions include whether or not a document was viewed or ignored and the time spent viewing the document. Documents to which the user responds positively are parsed and analyzed (200) and then used to update the User Model (202) as described above.

The role of the User Model in filtering the search results in step 194 is based on Bayesian statistics and pattern classification theory. According to pattern classification theory, as detailed in R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973, the optimal search result is the one with the highest posterior probability. That is, the optimal result is given by:

$$\text{Max}_d P(u|q, d),$$

where  $P(u|q, d)$  is the posterior probability of the event that a document  $d$  is of interest to a user  $u$  having an information need  $q$ . This probability can be expressed as:

$$P(u|q, d) = \frac{P(q|d, u) P(u|d)}{P(q|d)}.$$

The term  $P(u|d)$  represents the user interest in the document regardless of the current information need, and is calculated using the User Model. The term  $P(q|d, u)$  represents the probability that a user  $u$  with an information need of  $d$  expresses it in the form of a query  $q$ . The term  $P(q|d)$  represents the probability that an average user with an information need of  $d$  expresses it in the form of a query  $q$ . One possible implementation of the latter two terms uses the Hidden Markov Model, described in Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999.